

A *Tm4sf1*-Marked Subpopulation of Endothelial Stem/Progenitor Cells Identified by Lung Single-Cell Omics of Pulmonary Arterial Hypertension

Jason Hong^{1*}, Brenda Wong¹, Caroline Huynh², Brian Tang², Soban Umar³, Xia Yang^{2**‡},
Mansoureh Eghbali^{3**‡}

¹Division of Pulmonary and Critical Care Medicine, ²Department of Integrative Biology and Physiology, and ³Department of Anesthesiology & Perioperative Medicine, University of California, Los Angeles, Los Angeles, CA

*Corresponding authors

‡Co-senior authors.

Correspondence

Jason Hong, M.D., Ph.D.

Division of Pulmonary and Critical Care Medicine

David Geffen School of Medicine at UCLA

200 UCLA Medical Plaza, Suite 365-B, Box 951693

Los Angeles, CA 90095

E-mail: jasonhong@mednet.ucla.edu

Author Contributions: J.H. contributed to the conception and design of the research and interpretation of the data. J.H. conducted or helped with all experiments, analyzed the data, made the figures, and wrote the manuscript. B.W. helped with data interpretation, preparing figures

and writing the manuscript. S.U. performed the animal experiments and generated the bulk RNA-seq data. C.H. performed trajectory analysis. B.T. performed pathway enrichment analysis. X.Y. and M.E. provided intellectual input.

Abstract

Rationale: The identification and role of endothelial progenitor cells (EPCs) in pulmonary arterial hypertension (PAH) remains controversial. Single-cell omics analysis can shed light on EPCs and their potential contribution to PAH pathobiology.

Objectives: We aim to identify EPCs in rat lungs and assess their relevance to preclinical and human PAH.

Methods: Differential expression, gene set enrichment, cell-cell communication, and trajectory reconstruction analyses were performed on lung endothelial cells from single-cell RNA-seq of Sugen-hypoxia, monocrotaline, and control rats. Relevance to human PAH was assessed in multiple independent blood and lung transcriptomic datasets.

Measurements and Main Results: A subpopulation of endothelial cells (EA2) marked by *Tm4sf1*, a gene strongly implicated in cancer, harbored a distinct transcriptomic signature including *Bmpr2* downregulation that was enriched for pathways such as inflammation and angiogenesis. Cell-to-cell communication networks specific to EA2 were activated in PAH such as CXCL12 signaling. Trajectory analysis demonstrated EA2 has a stem/progenitor cell phenotype. Analysis of independent datasets revealed *Tm4sf1* is a marker for hematopoietic stem

cells and is upregulated in PAH peripheral blood, particularly in patients with worse WHO functional class. EA2 signature genes including *Procr* and *Sulf1* were found to be differentially regulated in the lungs of PAH patients and in PAH models *in vitro*, such as BMPR2 knockdown.

Conclusions: Our study uncovered a novel *Tm4sf1*-marked stem/progenitor subpopulation of rat lung endothelial cells and demonstrated its relevance to preclinical and human PAH. Future experimental studies are warranted to further elucidate the pathogenic role and therapeutic potential of targeting EA2 and *Tm4sf1* in PAH.

Abstract word count: 252

Key words: pulmonary hypertension; scRNA-seq; endothelial progenitor cells; monocrotaline; Sugen-hypoxia

Introduction

Despite advances in our understanding of pulmonary arterial hypertension (PAH), it remains an incurable disease characterized by pathological pulmonary arterial remodeling and endothelial dysfunction. Endothelial progenitor cells (EPCs) are involved in endothelial homeostasis and angiogenesis and have been extensively studied in PAH animal models (1–3) and patients (4–10) using heterogenous approaches for EPC isolation by flow cytometry and/or culture. However, results have been conflicting and potentially confounding as to whether EPCs in PAH are protective vs. harmful (11), a reflection of the broader controversy in the stem cell field as to the precise definition and reliable isolation of EPCs (12). For example, conventional markers used to identify putative EPCs in numerous papers including PAH studies are now believed to be

unreliable in isolating pure EPCs (12) and a specific marker for EPCs has yet to be identified (13).

The advent of single-cell omics can shed light on EPCs and their potential role in disease in a data-driven and unbiased approach. But while single-cell studies in PAH have recently been published including our recent work using lungs from PAH rat models (14–17), stem/progenitor endothelial cells have yet to be identified. However, rapidly advancing computational methods to dissect single-cell data provide an opportunity to unravel important biological discoveries. In this study of rat lung endothelial cells from single-cell RNA-sequencing (scRNA-seq), we uncover a *Tm4sf1*-marked subpopulation of endothelial stem/progenitor cells and demonstrate its relevance to PAH through its distinct transcriptomic signature, cell-to-cell signaling network, and differential regulation of its marker genes in circulating blood and lungs of PAH patients and animal models.

Methods

Main methods are below with additional details provided in a supplement.

Animals

Adult male Sprague-Dawley rats (250g-350g) were used for animal experiments, which were approved by the UCLA Animal Research Committee. As previously described (17), Sugen-hypoxia (SuHx) rats were injected subcutaneously with Sugen 5416 (20 mg/kg) followed by hypoxia at 10% O₂ for 21 days and then normoxia for 14 days. Monocrotaline (MCT) rats were injected subcutaneously with MCT (60 mg/kg) followed by normoxia for 28 days. Age-matched

control rats were kept in normoxia for 28 days. Lungs were then harvested and enzymatically dissociated into single-cell suspensions followed by scRNA-seq (17, 18) ($n = 6/\text{group}$) or snap-frozen for RNA isolation and bulk RNA-seq ($n = 4/\text{group}$).

Differential expression analysis

Expression data was normalized, filtered, and clustered using Seurat R package (13)(19). Cell types were identified based on known cell type marker genes and endothelial clusters were then used for all downstream analyses. Differentially expressed genes (DEGs) were determined using MAST (20). To annotate DEGs, Gene Set Enrichment Analysis (GSEA) was performed using R package fgsea v1.18.0. PAH genes were retrieved from DisGeNET (21) and Comparative Toxicogenomics Database (CTD) (22).

Cell-cell communication analysis

CellPhoneDB, a repository of ligands and receptors, was used to infer cell–cell communication from combined expression of multi-subunit ligand–receptor complexes in our scRNA-seq data (23). Enriched ligand–receptor interactions between two cell clusters are derived on the basis of expression of a receptor by one cell cluster and a ligand by another cell cluster. For each gene in the cluster, the percentage of cells expressing the gene and the gene expression mean are calculated. The expression levels of ligands and receptors within each cell cluster are considered and random shuffling is used to determine which ligand–receptor pairs display significant specificity between two cell clusters.

Trajectory reconstruction

To predict differentiation states, we used CytoTRACE, a computational framework that leverages single-cell gene counts as a determinant of developmental potential, covariant gene expression, and local neighborhoods of transcriptionally similar cells to predict ordered differentiation states from scRNA-seq data (24).

Independent datasets for validation

Select genes were queried in independent bulk and single-cell transcriptomic datasets of human lung and blood samples from Gene Expression Omnibus (GEO) or hosted on lab web servers.

Results

Tm4sf1-marked subpopulation of endothelial cells expresses a distinct transcriptomic signature relevant to PAH

The transcriptomes of 758 endothelial cells generated by scRNA-seq of 18 lungs from MCT, SuHx, and control rats (6/group) were clustered while retaining cell type annotations as previously described (17) yielding 3 distinct subclusters: 454 endothelial arterial type 1 cells (EA1), 255 endothelial arterial type 2 cells (EA2), and 49 endothelial capillary cells (Figure 1A). Each of the 3 subpopulations were represented in MCT, SuHx, and control lungs (Figure 1B). Subcluster-specific markers were identified such as *Nostrin* for EA1, *Tm4sf1* for EA2, and *Car4*, a known marker for EC (Figures 1C-1E). Arterial markers such as *Sox17* (Figure 1F) and *Efnb2* (Figure E1) were also identified confirming the arterial origin of EA1 and EA2 (25).

Given that the pathological hallmark of vascular remodeling in PAH predominantly occurs in the pulmonary arteries, we then evaluated the differences in gene expression between the two arterial subclusters EA1 and EA2. We found a total of 196 genes differentially expressed (DEGs; FDR < 0.05) with 117 genes upregulated and 79 genes downregulated in EA2 compared to EA1 (Figure 1G and Table E1). *Tm4sf1* was the top upregulated DEG in EA2 vs. EA1 and its expression was specific to EA2 when comparing all lung cell types (Figure E2). *Tm4sf1* encodes the cell-surface protein transmembrane 4 L six family member 1 which plays a role in the regulation of cell development, growth and motility and has been strongly implicated in various cancers (26) but has never before been studied in PAH. Given the importance of cancer-related processes and their overlap with PAH (27), we paid closer attention to the *Tm4sf1*-marked subpopulation of endothelial cells, EA2. *Nostrin* was the top downregulated gene in EA2 vs. EA1- it encodes nitric oxide synthase trafficking, an adaptor protein that binds to endothelial nitric oxide synthase (eNOS) and attenuates NO production. When intersecting other DEGs with disease-gene databases DisGeNET and CTD, we found that a number of DEGs are also known PAH genes such as *Bmpr2* which was downregulated in EA2 (Figure 1G).

We then used gene set enrichment analysis to determine what biological processes, pathways, cell types, and diseases are enriched in the gene signature of EA2 (vs. EA1). Among 5,938 biological processes from Gene Ontology (GO) (28), cytoplasmic translation and response to tumor necrosis factor (TNF) were most enriched in upregulated EA2 genes. Furthermore, angiogenesis-related processes were most enriched in downregulated EA2 genes (Figure 1H). Of the 50 Hallmark pathways (29), many known to be involved in PAH were enriched in genes upregulated in EA2 such as TNF α /NF- κ B signaling, apoptosis, hypoxia, p53 pathway, and

epithelial mesenchymal transition (EMT) (Figure 1I). To determine what cell type signatures may be enriched in EA2 vs. EA1 DEGs, we tested 344 cell type-specific gene sets from HubMAP (30) and unexpectedly found that immune cell signatures, particularly from monocytes, were most enriched in genes differentially upregulated in EA2 whereas endothelial signatures were most enriched in genes upregulated in EA1 (Figure 1J). Additionally, we found that across 9,602 disease-specific gene sets from DisGeNET, genes downregulated in EA2 vs. EA1 were most enriched for diseases associated with primary or secondary forms of pulmonary hypertension (PH): hereditary hemorrhagic telangiectasia (HHT), pulmonary veno-occlusive disease (PVOD), chronic thromboembolic pulmonary hypertension (CTEPH), and familial PAH (Figure 1K). Thus, not only was *Bmpr2* downregulated in EA2 as it is in human PAH lungs (31), the downregulated portion of the EA2 signature on a transcriptome-wide scale was also highly associated with PAH.

EA2-specific cell-to-cell signaling networks are activated in PAH

Having identified a transcriptionally distinct subpopulation of lung endothelial cells relevant to PAH, we next hypothesized that communication between EA2 and other cell types occurs in a disease-specific manner given that many other cell types besides endothelial cells have also been implicated in the complex pathobiology of PAH. Using CellPhoneDB to predict ligand-receptor interactions between cell types, we found overall increased intercellular communication between various immune, mesenchymal, and epithelial cells in the lungs of PAH rat models compared to control (Figure 2A). When quantifying the number of statistically significant ligand-receptor interactions between endothelial cells and other cell types, EA2 exhibited stronger cell-to-cell

signaling in MCT and SuHx lungs compared to control that was not seen in EA1 nor EC (Figure 2B).

We further examined specific ligand-receptor interactions and found that signaling molecules known to play a role in PAH such as vascular endothelial growth factor A (VEGFA) (32), C-X-C motif chemokine ligand 12 (CXCL12) (33, 34), and macrophage migration inhibitory factor (MIF) (35, 36) were not only differentially active in PAH models compared to control but also specific to cell-cell communication involving EA2 but not EA1 (Figure 2C) nor EC (Figure E3). For instance, VEGFA from cell types such as fibroblasts and interstitial macrophages were predicted to interact in a PAH- and EA2-specific manner with its cognate receptors fms related receptor tyrosine kinase 1 (FLT1; VEGFR-1) and kinase insert domain receptor (KDR; VEGFR-2). CXCL12 signaling was highly specific to EA2 with distinct cell-cell communication profiles specific to the particular receptor expressed on target cell types. For example, EA2's CXCL12 interaction with C-X-C motif chemokine receptor 4 (CXCR4) was observed in various PAH immune cell types of both lymphoid and myeloid lineage whereas its interaction with atypical chemokine receptor 3 (ACKR3; CXCR7) and dipeptidyl peptidase 4 (DPP4) was noted in PAH fibroblasts, endothelial, and mesothelial cells (Figure 2C). MIF is a key pro-inflammatory immune regulator whose interaction with its receptor CD74 on myeloid and B cells was statistically significant for both MCT and SuHx EA2 compared to control, a pattern that was unique to EA2 (Figure 2C).

Trajectory analysis reveals that EA2 has a stem/progenitor cell phenotype

Given the important role of CXCL12 and its chemokine receptor CXCR4 in the homing and function of endothelial progenitor cells (EPCs) (9, 37–40) and the specificity of this signaling axis for EA2 communication in our rat lung scRNA-seq analysis, we asked whether EA2 might have a stem/progenitor cell phenotype. We used CytoTRACE (24), an unbiased computational method that leverages the number of expressed genes per cell as a determinant of developmental potential to predict the differentiation state of single cells from our endothelial scRNA-seq data (Figure 3A). Compared to EA1 and EC cells, we found that EA2 cells had significantly higher CytoTRACE scores corresponding to a less differentiated state (Figure 3B). Comparing CytoTRACE scores by condition, endothelial cells from MCT and SuHx lungs were predicted to be less differentiated compared to control, consistent with prior evidence implicating endothelial progenitor cells in PAH (9, 41, 42). To identify markers of the less differentiated phenotype, all 12,116 genes with detectable expression in our dataset were rank-ordered on the basis of their correlation with CytoTRACE scores across all 758 endothelial cells analyzed. Ribosomal genes correlated most strongly with a less differentiated state (Figure 3D) and thus were downregulated with a more differentiated state, consistent with prior studies (43, 44). Gene set enrichment analysis revealed many GO biological processes significantly enriched in genes correlated with a less differentiated state compared to a more differentiated state (219 vs. 0 out of 5,994 total gene sets) (Figure E4). Top enriched biological processes included cotranslational protein targeting to membrane and cytoplasmic translation which were also top enriched GO biological processes of the EA2 gene signature (Figure 1H) providing further evidence for a stem/progenitor phenotype of EA2.

Tm4sf1 is a marker for hematopoietic stem cells and is upregulated in the peripheral blood of PAH patients

Given prior studies implicating circulating EPCs in PAH patients (6–9), we asked whether *Tm4sf1*, the most highly specific marker for rat lung EA2, could also be a marker for circulating stem/progenitor cells in human PAH. Using a dataset of 7,551 human blood cell transcriptomes representing 43 cell type clusters derived from 21 healthy donors (45), we found that *TM4SF1* was specifically expressed in and one of 19 marker genes for hematopoietic stem cell/multipotent progenitor 1 cells (HSC/MPP1) (FDR = 2.0e-15) (Figure 4A) further suggesting that *Tm4sf1*-marked EA2 cells may be a stem/progenitor subpopulation of endothelial cells. Supporting the relevance of this marker to PAH, we found in two independent microarray datasets that *TM4SF1* was significantly upregulated in circulating peripheral blood mononuclear cells (PBMCs) of PAH patients compared to healthy control ($n = 72$ and 41 , respectively) (46) and in systemic sclerosis (SSc)-associated PAH patients compared to SSc without PAH ($n = 10$ and 10 , respectively) (47) (Figures 4B-4C). Furthermore, in a third dataset of whole blood RNA-seq, PAH patients with higher *TM4SF1* expression were associated with worse WHO functional class (FC) ($n = 65$, 239 , and 34 for FC II, III, and IV, respectively) (Figure 4D).

EA2 signature genes are differentially regulated in the lungs of PAH patients and in PAH models of human endothelial cells

Having established that the EA2 marker *Tm4sf1* is upregulated in PAH peripheral blood and associated with disease severity, we next investigated whether *Tm4sf1* and other EA2 markers are dysregulated in PAH lungs. Given that the power to detect PAH-specific differences in *Tm4sf1* expression in our rat lung scRNA-seq dataset may have been limited by relatively lower

yield of endothelial cells from the tissue dissociation into single cells, we evaluated whether *Tm4sf1* expression is dysregulated at the bulk lung tissue level. We performed bulk RNA sequencing on lung tissue from a subset of the same animals used for scRNA-seq ($n = 4/\text{group}$) and found that *Tm4sf1* was significantly upregulated in the lungs of both MCT and SuHx compared to control (Figure 5A). Supporting the human relevance of this finding, a 2006 study from Harefield Hospital in the United Kingdom which used suppression subtractive hybridization found that *TM4SF1* was one of 27 upregulated genes in PAH vs. donor lungs ($n = 2/\text{group}$) (48). However, we probed a 2010 University of Pittsburgh (Pitt) microarray dataset (49) and found that *TM4SF1* was instead downregulated in PAH lungs compared to control ($n = 18$ and 13, respectively) (Figure 5B) which could reflect differences in patient population and disease severity between datasets. Furthermore, analysis of a 2020 Pitt scRNA-seq dataset revealed that *TM4SF1* was downregulated in endothelial cells from PAH lungs compared to donors ($n = 3$ and 6, respectively) (15) (Figure 5C). We then evaluated two other top EA2 markers *Procr* and *Sulf1* and discovered both to be upregulated in the lungs of PAH patients in two independent microarray datasets each (Figures 5D-5G) (49–51).

Having determined that top EA2 markers are dysregulated in the lungs of PAH patients and given that *Bmpr2*, the most well-established causal gene in PAH, is downregulated in EA2 compared to EA1 (Figure 1G), we next asked whether EA2 markers are regulated by *Bmpr2*. We analyzed a microarray of siRNA knockdown of *BMPR2* in primary human pulmonary artery endothelial cells (PAECs) *in vitro* (52) and found no change in *TM4SF1* but *PROCR* and *SULF1* were upregulated consistent with upregulation in EA2 vs. EA1 and PAH vs. donor lungs (Figures

5H-5K). Furthermore, *NOSTRIN*, the top downregulated gene in EA2 vs. EA1 (Figure 1G), was also downregulated in human PAECs upon *BMPR2* knockdown (Figure 5K).

Given the strong connection between *Tm4sf1* and various cancers (26), we then evaluated the response of *Tm4sf1* and other signature genes of EA2 to phorbol 12-myristate 13-acetate (PMA), a potent tumor promoter which activates protein kinase C and downstream mitogen-activated protein kinase (MAPK) pathways which have been implicated in PAH pathogenesis (52–54). We found that *TM4SF1* was upregulated in *BMPR2*-silenced human PAECs exposed to PMA (Figure 5L), while *PROCR*, *SULF1*, and *NOSTRIN* were either downregulated or unchanged (Figures 5M-5O). Overall, we found significant correlation between DEGs in EA2 (vs. EA1) and DEGs in human PAECs after either *BMPR2* knockdown (Figure 5P) or PMA stimulation (Figure 5Q) suggesting that portions of the EA2 signature can be explained by both downregulation of *Bmpr2* and stimulation with a tumor promoter. Overlapping DEGs were enriched for known PAH genes (Figures 5P-5Q) and pathways (Figures E6A-E6B) in both conditions, such as EMT in *BMPR2*-silenced PAECs and TNF α /NF- κ B signaling in PMA-simulated PAECs.

Discussion

In this study, we uncover a *Tm4sf1*-marked subpopulation of rat lung endothelial cells and demonstrate its relevance to PAH through its distinct transcriptomic signature, cell-to-cell signaling network, stem/progenitor cell phenotype, and differential regulation of its marker genes in circulating blood and lungs of PAH patients and animal models.

Tm4sf1, the top marker of EA2, encodes a cell-surface protein and member of the transmembrane 4 superfamily which mediates signal transduction events that regulate cell development, activation, growth, and motility(55). TM4SF1 is highly expressed and associated with poor survival in a wide range of cancers (26) but has never previously been studied in PAH, a disease that shares many pathobiological features with cancer for which many antineoplastic therapies have been evaluated for potential repurposing in PAH (27). With the accumulating evidence for TM4SF1's role in cancer, potential oncologic therapies have been investigated such as in a phase 1/2 clinical trial evaluating chimeric antigen receptor T-cells (CAR-T) to treat patients with TM4SF1-positive pancreatic, colorectal, gastric, and lung cancer (ClinicalTrials NCT04151186). TM4SF1 promotes the migration and invasion of cancer cells by inducing EMT, self-renewal ability, and tumor angiogenesis (26), all of which are mechanisms relevant to the pathological pulmonary vascular remodeling in PAH.

Similar to malignant tumors, we found *Tm4sf1* to be upregulated in the lungs of SuHx and MCT rats (at the standard time points of 5 total weeks and 4 weeks, respectively) and in the lungs of PAH patients undergoing transplant from a 2006 UK study. However, *TM4SF1* was downregulated at the lung tissue level in a 2010 Pitt microarray and at the lung endothelial cell level in a 2020 Pitt scRNA-seq dataset in explanted PAH lungs. Differences in the directionality of *TM4SF1*'s dysregulation in PAH could reflect institutional, national, and/or temporal differences in the characteristics of patients undergoing lung transplant such as disease stage, severity, and/or treatment history. We hypothesize that TM4SF1 is upregulated earlier in the disease course and then becomes downregulated in end-stage PAH, both of which may be harmful. Supporting this hypothesis, knockdown of TM4SF1 in PAECs causes endothelial

senescence (56) which has recently been shown to be causal in the transition from a reversible to the irreversible pulmonary vascular phenotype of end-stage PAH (57).

Furthermore, we found that not only was *TM4SF1* dysregulated in PAH lungs, it was upregulated in PAH peripheral blood in multiple independent cohorts, associated with disease severity, and a highly specific marker for HSCs, which reside in bone marrow, circulate in peripheral blood, and give rise to not only all blood cell lineages, but also EPCs (58). Given that *Tm4sf1* was also a highly specific marker for EA2 in our rat lung dataset, this raises the possibility that EA2 is a stem/progenitor subpopulation of lung endothelial cells that derives from bone marrow via the circulating blood. Consistent with a stem/progenitor phenotype, our trajectory analysis also predicted that EA2 cells are of an earlier differentiation state compared to other endothelial cells. Supporting this finding, a recent study identified *Tm4sf1* as a marker of multipotent endothelial precursor cells by scRNA-seq of lineage-traced mouse lungs (59).

While controversy and uncertainty persists in the definition and isolation of EPCs and their role in PAH (11–13, 60), it is generally accepted there exists 2 distinct subsets: so-called late EPCs derived from endothelial lineage and early EPCs derived from hematopoietic lineage with which EA2 shares features. Not only does EA2 share a novel marker with HSCs supporting its hematopoietic origin, we also unexpectedly found that the EA2 signature is highly enriched for myeloid signatures (Figure 1J) which is similarly noted in early EPCs (61). In fact, renaming of early EPCs to “myeloid angiogenic cells” has been proposed to more precisely define its phenotype and function (12) such as secretion of key proangiogenic factors like CXCL12 (60) - a chemokine implicated in PAH (33, 34, 62, 63) whose interaction with various target cells were

unique to EA2 (Figure 2C). Consistent with our finding that EA2's marker *TM4SF1* is upregulated in PAH circulation, a previous study found that early EPCs, but not late EPCs, are increased in PAH circulation as well (8). In other studies of early EPCs, one found a trend towards an increase in PAH (7), and the other found a decrease in PAH (6) but used an older isolation method now believed to be unreliable in EPC purification (13).

Further indicating EA2's relevance in PAH pathobiology, we also demonstrated that two other top EA2 markers, *Sulfl* and *Procr*, were upregulated in explanted PAH lungs in multiple independent cohorts. In line with our findings, SULF1 (Sulfatase-1) was recently shown to be upregulated in remodeled pulmonary arterioles in PAH patients and MCT and SuHx rats (64). Further supporting EA2's stem/progenitor phenotype, *Procr*, encoding protein C receptor, was previously identified as a marker for murine HSCs (65) and blood vascular endothelial stem cells (66) which displayed EMT and angiogenic signatures. While upregulated EA2 genes (relative to EA1) were enriched for EMT, genes downregulated in EA2 were enriched for angiogenesis (Figures 1H-1I). For example, the top downregulated gene in EA2, NOSTRIN (nitric oxide synthase trafficking), is known to inhibit angiogenesis (67) thus supporting an angiogenic phenotype of EA2. Furthermore, NOSTRIN has also been shown to suppress TNF α /NF- κ B signaling (67) which is congruent with TNF α /NF- κ B signaling as the top Hallmark pathway upregulated in the EA2 signature (Figure 1I).

Another downregulated gene in EA2 was *Bmpr2*, the most well-established causal PAH gene whose expression is also decreased in patient lungs of various PAH subtypes including non *BMPR2* mutation-related PAH (31). Reduced *BMPR2* expression is known to cause exaggerated

TGF-beta signaling in line with our finding that TGF-beta signaling was the top Hallmark pathway enriched in genes downregulated in EA2 (Figure 1I). Overactive TGF-beta signaling ultimately leads to pulmonary vascular remodeling in PAH through processes such as inflammation, angiogenesis, and EndMT (68), all of which are also upregulated in EA2. Furthermore, we demonstrated that *Sulf1*, *Procr*, *Nostrin* and a number of other genes are all differentially regulated after *BMPR2* knockdown in human PAECs in the same direction as EA2's signature suggesting these genes may be downstream of *BMPR2* signaling and thus downregulation of *Bmpr2* likely induces a portion of EA2's gene signature. *TM4SF1* was unchanged upon *BMPR2* knockdown suggesting it may act upstream or independently of *BMPR2* signaling, or a difference was not detected due to variability in experimental replicates. However, *TM4SF1* and many other EA2 signature genes including known PAH genes were upregulated upon PMA stimulation suggesting their potential regulation by MAPK signaling pathways which have been implicated in PAH pathogenesis (52–54).

Given the challenges of enzymatically dissociating tissue into single cells that may be particularly fragile or tightly embedded, we cannot exclude the possibility that other arterial endothelial populations were not captured in our scRNA-seq dataset. Alternative methods such as single-nucleus RNA-seq may provide a more precise cellular representation of tissue composition (69).

In conclusion, our study provides an in-depth analysis of rat lung endothelial heterogeneity using unbiased single-cell computational methods to uncover a novel *Tm4sf1*-marked stem/progenitor subpopulation with high relevance to PAH supported by multiple independent human and rat

datasets. Future experimental studies are warranted to further elucidate the pathogenic role and therapeutic potential of targeting EA2 and *Tm4sf1* in PAH.

Acknowledgements

The authors would like to express gratitude to Dr. Douglas Arneson for his guidance with scRNA-seq analytical techniques, Drs. In Sook Ahn and Graciél Diamante for their help in performing Drop-seq, Drs. Christine Cunningham and May Bhetraratana for their contribution to lung tissue dissociation, and Dr. Gregoire Ruffenach for his assistance with the animal experiments.

References

1. Nagaya N, Kangawa K, Kanda M, Uematsu M, Horio T, Fukuyama N, Hino J, Harada-Shiba M, Okumura H, Tabata Y, Mochizuki N, Chiba Y, Nishioka K, Miyatake K, Asahara T, Hara H, Mori H. Hybrid Cell–Gene Therapy for Pulmonary Hypertension Based on Phagocytosing Action of Endothelial Progenitor Cells. *Circulation* 2003;108:889–895.
2. Zhao YD. Rescue of Monocrotaline-Induced Pulmonary Arterial Hypertension Using Bone Marrow-Derived Endothelial-Like Progenitor Cells: Efficacy of Combined Cell and eNOS Gene Therapy in Established Disease. *Circulation Research* 2005;96:442–450.
3. Harper RL, Maiolo S, Ward RJ, Seyfang J, Cockshell MP, Bonder CS, Reynolds PN. BMPR2-expressing bone marrow-derived endothelial-like progenitor cells alleviate pulmonary arterial hypertension in vivo. *Respirology* 2019;24:1095–1103.

4. Wang X-X, Zhang F-R, Shang Y-P, Zhu J-H, Xie X-D, Tao Q-M, Zhu J-H, Chen J-Z. Transplantation of Autologous Endothelial Progenitor Cells May Be Beneficial in Patients With Idiopathic Pulmonary Arterial Hypertension. *Journal of the American College of Cardiology* 2007;49:1566–1571.
5. Zhu JH, Wang XX, Zhang FR, Shang YP, Tao QM, Zhu JH, Chen JZ. Safety and efficacy of autologous endothelial progenitor cells transplantation in children with idiopathic pulmonary arterial hypertension: open-label pilot study. *Pediatr Transplant* 2008;12:650–655.
6. Junhui Z, Xingxiang W, Guosheng F, Yunpeng S, Furong Z, Junzhu C. Reduced number and activity of circulating endothelial progenitor cells in patients with idiopathic pulmonary arterial hypertension. *Respir Med* 2008;102:1073–1079.
7. Diller G-P, van Eijl S, Okonko DO, Howard LS, Ali O, Thum T, Wort SJ, Bédard E, Gibbs JSR, Bauersachs J, Hobbs AJ, Wilkins MR, Gatzoulis MA, Wharton J. Circulating Endothelial Progenitor Cells in Patients With Eisenmenger Syndrome and Idiopathic Pulmonary Arterial Hypertension. *Circulation* 2008;117:3020–3030.
8. Asosingh K, Aldred MA, Vasanji A, Drazba J, Sharp J, Farver C, Comhair SAA, Xu W, Licina L, Huang L, Anand-Apte B, Yoder MC, Tudor RM, Erzurum SC. Circulating angiogenic precursors in idiopathic pulmonary arterial hypertension. *Am J Pathol* 2008;172:615–627.
9. Toshner M, Voswinckel R, Southwood M, Al-Lamki R, Howard LSG, Marchesan D, Yang J, Suntharalingam J, Soon E, Exley A, Stewart S, Hecker M, Zhu Z, Gehling U, Seeger W, Pepke-Zaba J, Morrell NW. Evidence of Dysfunction of Endothelial Progenitors in Pulmonary Arterial Hypertension. *Am J Respir Crit Care Med* 2009;180:780–787.

10. Granton J, Langleben D, Kutryk MB, Camack N, Galipeau J, Courtman DW, Stewart DJ. Endothelial NO-Synthase Gene-Enhanced Progenitor Cell Therapy for Pulmonary Arterial Hypertension: The PHACeT Trial. *Circ Res* 2015;117:645–654.
11. Fadini GP, Avogaro A, Ferraccioli G, Agostini C. Endothelial progenitors in pulmonary hypertension: new pathophysiology and therapeutic implications. *European Respiratory Journal* 2010;35:418–425.
12. Medina RJ, Barber CL, Sabatier F, Dignat-George F, Melero-Martin JM, Khosrotehrani K, Ohneda O, Randi AM, Chan JKY, Yamaguchi T, Van Hinsbergh VWM, Yoder MC, Stitt AW. Endothelial Progenitors: A Consensus Statement on Nomenclature. *Stem Cells Transl Med* 2017;6:1316–1320.
13. Yoder MC. Human Endothelial Progenitor Cells. *Cold Spring Harb Perspect Med* 2012;2:a006692.
14. Asosingh K, Comhair S, Mavrikakis L, Xu W, Horton D, Taylor I, Tkachenko S, Hu B, Erzurum S. Single-cell transcriptomic profile of human pulmonary artery endothelial cells in health and pulmonary arterial hypertension. *Sci Rep* 2021;11:14714.
15. Saygin D, Tabib T, Bittar HET, Valenzi E, Sembrat J, Chan SY, Rojas M, Lafyatis R. Transcriptional profiling of lung cell populations in idiopathic pulmonary arterial hypertension. *Pulm Circ* 2020;10:.
16. Rodor J, Chen S-H, Scanlon JP, Monteiro JP, Caudrillier A, Sweta S, Stewart KR, Shmakova A, Dobie R, Henderson BEP, Stewart K, Hadoke PWF, Southwood M, Moore SD, Upton PD, Morrell NW, Li Z, Chan SY, Handen A, Lafyatis R, de Rooij LPMH, Henderson NC, Carmeliet P, Spiroski A-M, Brittan M, Baker AH. Single-cell RNA sequencing profiling of

- mouse endothelial cells in response to pulmonary arterial hypertension. *Cardiovascular Research* 2021;cvab296.doi:10.1093/cvr/cvab296.
17. Hong J, Arneson D, Umar S, Ruffenach G, Cunningham CM, Ahn IS, Diamante G, Bhetraratana M, Park JF, Said E, Huynh C, Le T, Medzikovic L, Humbert M, Soubrier F, Montani D, Girerd B, Trégouët D-A, Channick R, Saggar R, Eghbali M, Yang X. Single-Cell Study of Two Rat Models of Pulmonary Arterial Hypertension Reveals Connections to Human Pathobiology and Drug Repositioning. *Am J Respir Crit Care Med* 2021;203:1006–1022.
 18. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 2015;161:1202–1214.
 19. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* 2018;36:411–420.
 20. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, Linsley PS, Gottardo R. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 2015;16:.
 21. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 2017;45:D833–D839.

22. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res* 2019;47:D948–D954.
23. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* 2020;15:1484–1506.
24. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, Ilagan F, Kuo AH, Hsieh RW, Cai S, Zabala M, Scheeren FA, Lobo NA, Qian D, Yu FB, Dirbas FM, Clarke MF, Newman AM. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 2020;367:405–411.
25. Schupp JC, Adams TS, Cosme C, Raredon MSB, Yuan Y, Omote N, Poli S, Chioccioli M, Rose K-A, Manning EP, Sauler M, DeIuliis G, Ahangari F, Neumark N, Habermann AC, Gutierrez AJ, Bui LT, Lafyatis R, Pierce RW, Meyer KB, Nawijn MC, Teichmann SA, Banovich NE, Kropski JA, Niklason LE, Pe’er D, Yan X, Homer RJ, Rosas IO, *et al.* Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung. *Circulation* 2021;144:286–302.
26. Fu F, Yang X, Zheng M, Zhao Q, Zhang K, Li Z, Zhang H, Zhang S. Role of Transmembrane 4 L Six Family 1 in the Development and Progression of Cancer. *Frontiers in Molecular Biosciences* 2020;7:202.
27. Boucherat O, Vitry G, Trinh I, Paulin R, Provencher S, Bonnet S. The cancer theory of pulmonary arterial hypertension. *Pulmonary Circulation* 2017;7:285–299.
28. The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, Basu S, Chisholm RL, Dodson RJ, Hartline E, Fey P, Thomas PD, Albou L-P,

- Ebert D, Kesling MJ, Mi H, Muruganujan A, Huang X, Mushayahama T, LaBonte SA, Siegele DA, Antonazzo G, Attrill H, Brown NH, Garapati P, Marygold SJ, Trovisco V, dos Santos G, *et al.* The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* 2021;49:D325–D334.
29. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–425.
30. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* 2019;574:187–192.
31. Atkinson C. Primary Pulmonary Hypertension Is Associated With Reduced Pulmonary Vascular Expression of Type II Bone Morphogenetic Protein Receptor. *Circulation* 2002;105:1672–1678.
32. Voelkel NF, Gomez-Arroyo J. The role of vascular endothelial growth factor in pulmonary arterial hypertension. The angiogenesis paradox. *Am J Respir Cell Mol Biol* 2014;51:474–484.
33. Bordenave J, Thuillet R, Tu L, Phan C, Cumont A, Marsol C, Huertas A, Savale L, Hibert M, Galzi J-L, Bonnet D, Humbert M, Frossard N, Guignabert C. Neutralization of CXCL12 attenuates established pulmonary hypertension in rats. *Cardiovasc Res* 2020;116:686–697.
34. McCullagh BN, Costello CM, Li L, O’Connell C, Codd M, Lawrie A, Morton A, Kiely DG, Condliffe R, Elliot C, McLoughlin P, Gaine S. Elevated Plasma CXCL12 α Is Associated with a Poorer Prognosis in Pulmonary Arterial Hypertension. *PLOS ONE* 2015;10:e0123709.
35. Le Hiress M, Tu L, Ricard N, Phan C, Thuillet R, Fadel E, Dorfmueller P, Montani D, de Man F, Humbert M, Huertas A, Guignabert C. Proinflammatory Signature of the Dysfunctional Endothelium in Pulmonary Hypertension. Role of the Macrophage Migration Inhibitory

- Factor/CD74 Complex. *American Journal of Respiratory and Critical Care Medicine* 2015;192:983–997.
36. Jalce G, Guignabert C. Multiple roles of macrophage migration inhibitory factor in pulmonary hypertension. *Am J Physiol Lung Cell Mol Physiol* 2020;318:L1–L9.
37. Yamaguchi J, Kusano KF, Masuo O, Kawamoto A, Silver M, Murasawa S, Bosch-Marce M, Masuda H, Losordo DW, Isner JM, Asahara T. Stromal Cell–Derived Factor-1 Effects on Ex Vivo Expanded Endothelial Progenitor Cell Recruitment for Ischemic Neovascularization. *Circulation* 2003;107:1322–1328.
38. De Falco E, Porcelli D, Torella AR, Straino S, Iachininoto MG, Orlandi A, Truffa S, Biglioli P, Napolitano M, Capogrossi MC, Pesce M. SDF-1 involvement in endothelial phenotype and ischemia-induced recruitment of bone marrow progenitor cells. *Blood* 2004;104:3472–3482.
39. Xiao Q, Ye S, Oberhollenzer F, Mayr A, Jahangiri M, Willeit J, Kiechl S, Xu Q. SDF1 Gene Variation Is Associated with Circulating SDF1 α Level and Endothelial Progenitor Cell Number–The Bruneck Study. *PLOS ONE* 2008;3:e4061.
40. Walter DH, Haendeler J, Reinhold J, Rochwalsky U, Seeger F, Honold J, Hoffmann J, Urbich C, Lehmann R, Arenzana-Seisdesdos F, Aicher A, Heeschen C, Fichtlscherer S, Zeiher AM, Dimmeler S. Impaired CXCR4 Signaling Contributes to the Reduced Neovascularization Capacity of Endothelial Progenitor Cells From Patients With Coronary Artery Disease. *Circulation Research* 2005;97:1142–1151.
41. Montani D, Perros F, Gambaryan N, Girerd B, Dorfmueller P, Price LC, Huertas A, Hammad H, Lambrecht B, Simonneau G, Launay J-M, Cohen-Kaminsky S, Humbert M. C-kit-

positive cells accumulate in remodeled vessels of idiopathic pulmonary arterial hypertension.

Am J Respir Crit Care Med 2011;184:116–123.

42. Liu Y, Zhao X, Ding J, Xing Y, Zhou M, Wang X, Zhu W, Huo L, Yang J. Evidence of Accumulated Endothelial Progenitor Cells in the Lungs of Rats with Pulmonary Arterial Hypertension by ⁸⁹Zr-oxine PET Imaging. *Molecular Therapy - Methods & Clinical Development* 2020;17:1108–1117.
43. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun* 2017;8:2045.
44. Hayashi Y, Kuroda T, Kishimoto H, Wang C, Iwama A, Kimura K. Downregulation of rRNA Transcription Triggers Cell Differentiation. *PLOS ONE* 2014;9:e98586.
45. Xie X, Liu M, Zhang Y, Wang B, Zhu C, Wang C, Li Q, Huo Y, Guo J, Xu C, Hu L, Pang A, Ma S, Wang L, Cao W, Chen S, Li Q, Zhang S, Zhao X, Zhou W, Luo H, Zheng G, Jiang E, Feng S, Chen L, Shi L, Cheng H, Hao S, Zhu P, *et al.* Single-cell transcriptomic landscape of human blood cells. *National Science Review* 2021;8:nwaa180.
46. Cheadle C, Berger AE, Mathai SC, Grigoryev DN, Watkins TN, Sugawara Y, Barkataki S, Fan J, Boorgula M, Hummers L, Zaiman AL, Girgis R, McDevitt MA, Johns RA, Wigley F, Barnes KC, Hassoun PM. Erythroid-Specific Transcriptional Changes in PBMCs from Pulmonary Hypertension Patients. *PLOS ONE* 2012;7:e34951.
47. Hsu E, Shi H, Jordan RM, Lyons-Weiler J, Pilewski JM, Feghali-Bostwick CA. Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension. *Arthritis & Rheumatism* 2011;63:783–794.

48. Edgar AJ, Chacón MR, Bishop AE, Yacoub MH, Polak JM. Upregulated Genes In Sporadic, Idiopathic Pulmonary Arterial Hypertension. *Respir Res* 2006;7:1.
49. Rajkumar R, Konishi K, Richards TJ, Ishizawar DC, Wiechert AC, Kaminski N, Ahmad F. Genomewide RNA expression profiling in lung identifies distinct signatures in idiopathic pulmonary arterial hypertension and secondary pulmonary hypertension. *AJP: Heart and Circulatory Physiology* 2010;298:H1235–H1248.
50. Stearman RS, Bui QM, Speyer G, Handen A, Cornelius AR, Graham BB, Kim S, Mickler EA, Tudor RM, Chan SY, Geraci MW. Systems Analysis of the Human Pulmonary Arterial Hypertension Lung Transcriptome. *Am J Respir Cell Mol Biol* 2019;60:637–649.
51. Zhao YD, Yun HZH, Peng J, Yin L, Chu L, Wu L, Michalek R, Liu M, Keshavjee S, Waddell T, Granton J, de Perrot M. De novo synthesize of bile acids in pulmonary arterial hypertension lung. *Metabolomics* 2014;10:1169–1175.
52. Awad KS, Elinoff JM, Wang S, Gairhe S, Ferreyra GA, Cai R, Sun J, Solomon MA, Danner RL. Raf/ERK drives the proliferative and invasive phenotype of BMPR2-silenced pulmonary artery endothelial cells. *Am J Physiol Lung Cell Mol Physiol* 2016;310:L187-201.
53. Church AC, Martin DH, Wadsworth R, Bryson G, Fisher AJ, Welsh DJ, Peacock AJ. The reversal of pulmonary vascular remodeling through inhibition of p38 MAPK-alpha: a potential novel anti-inflammatory strategy in pulmonary hypertension. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2015;309:L333–L347.
54. Shafiq M, Jagavelu K, Iqbal H, Yadav P, Chanda D, Verma NK, Ghosh JK, Gaestel M, Hanif K. Inhibition of Mitogen-Activated Protein Kinase (MAPK)-Activated Protein Kinase 2 (MK2) is Protective in Pulmonary Hypertension. *Hypertension* 2021;77:1248–1259.

55. Huang Y-K, Fan X-G, Qiu F. TM4SF1 Promotes Proliferation, Invasion, and Metastasis in Human Liver Cancer Cells. *Int J Mol Sci* 2016;17:661.
56. Shih S-C, Zukauskas A, Li D, Liu G, Ang L-H, Nagy JA, Brown LF, Dvorak HF. The L6 Protein TM4SF1 Is Critical for Endothelial Cell Function and Tumor Angiogenesis. *Cancer Res* 2009;69:3272–3277.
57. van der Feen DE, Bossers GPL, Hagdorn QAJ, Moonen J-R, Kurakula K, Szulcek R, Chappell J, Vallania F, Donato M, Kok K, Kohli JS, Petersen AH, van Leusden T, Demaria M, Goumans M-JTH, De Boer RA, Khatri P, Rabinovitch M, Berger RMF, Bartelds B. Cellular senescence impairs the reversibility of pulmonary arterial hypertension. *Sci Transl Med* 2020;12:eaaw4974.
58. Bailey AS, Jiang S, Afentoulis M, Baumann CI, Schroeder DA, Olson SB, Wong MH, Fleming WH. Transplanted adult hematopoietic stems cells differentiate into functional endothelial cells. *Blood* 2004;103:13–19.
59. 2020 Annual World Congress of the Pulmonary Vascular Research Institute. *Pulm Circ* 2021;11:2045894020976949.
60. Pu X, Du L, Hu Y, Fan Y, Xu Q. Stem/Progenitor Cells and Pulmonary Arterial Hypertension. *Arteriosclerosis, Thrombosis, and Vascular Biology* 2021;41:167–178.
61. Medina RJ, O'Neill CL, Sweeney M, Guduric-Fuchs J, Gardiner TA, Simpson DA, Stitt AW. Molecular analysis of endothelial progenitor cell (EPC) subtypes reveals two distinct cell populations with different identities. *BMC Med Genomics* 2010;3:18.
62. Yi D, Liu B, Wang T, Liao Q, Zhu MM, Zhao Y-Y, Dai Z. Endothelial Autocrine Signaling through CXCL12/CXCR4/FoxM1 Axis Contributes to Severe Pulmonary Arterial Hypertension. *Int J Mol Sci* 2021;22:3182.

63. Yuan K, Liu Y, Zhang Y, Nathan A, Tian W, Yu J, Sweatt AJ, Shamshou EA, Condon D, Chakraborty A, Agarwal S, Auer N, Zhang S, Wu JC, Zamanian RT, Nicolls MR, de Jesus Perez VA. Mural Cell SDF1 Signaling Is Associated with the Pathogenesis of Pulmonary Arterial Hypertension. *Am J Respir Cell Mol Biol* 2020;62:747–759.
64. Samokhin A, Yefang W, Wertheim BM, Alba GA, Arons E, Maron B. Abstract 12577: Pulmonary Endothelial Sulfatase-1 is Regulated by NEDD9 and Increased in Pulmonary Arterial Hypertension. *Circulation* 2021;144:A12577–A12577.
65. Balazs AB, Fabian AJ, Esmon CT, Mulligan RC. Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow. *Blood* 2006;107:2317–2321.
66. Yu QC, Song W, Wang D, Zeng YA. Identification of blood vascular endothelial stem cells by the expression of protein C receptor. *Cell Res* 2016;26:1079–1098.
67. Chakraborty S, Ain R. Nitric-oxide synthase trafficking inducer is a pleiotropic regulator of endothelial cell function and signaling. *J Biol Chem* 2017;292:6600–6620.
68. Rol N, Kurakula KB, Happé C, Bogaard HJ, Goumans M-J. TGF- β and BMPR2 Signaling in PAH: Two Black Sheep in One Family. *Int J Mol Sci* 2018;19:2585.
69. Wu H, Kirita Y, Donnelly EL, Humphreys BD. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J Am Soc Nephrol* 2019;30:23–32.

Figure Legends

Figure 1: *Tm4sf1*-marked subpopulation of endothelial cells expresses a distinct transcriptomic signature relevant to PAH. (A-F) Uniform manifold approximation and projection (UMAP) plots showing clustering of 758 endothelial cell transcriptomes from 18 rats lungs with individual cells colored by (A) subpopulation, (B) disease condition ($n = 6/\text{group}$), and expression levels of the following marker genes: (C) *Car4* for endothelial capillary (EC), (D) *Nostrin* for endothelial arterial type 1 (EA), (E) *Tm4sf1* for endothelial arterial type 2 (EA2), and (F) *Sox17* as a known arterial marker expressed in both EA1 and EA2. Expression levels are represented as log normalized counts. (G) Volcano plot showing DEGs in 255 EA2 cells compared to 454 EA1 cells where x-axis represents MAST z-scores and y-axis indicates $-\log_{10}(P)$. Significant upregulated ($z > 0$) or downregulated ($z < 0$) genes with $\text{FDR} < 0.05$ are shown as purple and green dots, respectively. Select top DEGs are labeled with their gene names. DEGs highlighted in yellow represent human PAH-associated genes from either (black text) or both (red text) CTD and DisGeNET databases. (H-K) Dots plots showing Gene Set Enrichment Analysis (GSEA) of the EA2 DEG signature using (H) GO, (I) Hallmark, (J) HuBMAP, and (K) DisGeNET gene sets where x-axis represents normalized enrichment scores (NES) in which NES greater than or less than zero and $\text{FDR} < 0.05$ represent gene sets significantly enriched in up- or downregulated genes, respectively and are colored in purple or green, respectively. The y-axis represents gene sets ordered by their NES. Select gene sets are labeled and numbered by their ordering as top gene sets enriched in up- or downregulated genes. Dots larger in size represent higher $-\log_{10}(\text{FDR})$ values. $\text{FDR} = \text{false discovery rate}$.

Figure 2: EA2-specific cell-to-cell signaling networks are activated in PAH. (A) Network visualizations of control, MCT, and SuHx lungs where nodes are represented as cell types and

edge weight is proportional to the number of distinct and statistically significant ligand-receptor interaction pairs between two cell types as determined by CellPhoneDB. Thicker edges correspond to a higher number of ligand-receptor pairs. Node size is proportional to the sum of edge weights representing a given cell type's interactions with all other cell types. The top 10% of edges by weight are shown for each condition. (B) Box plots comparing cell-cell interactions where each dot represents the number of statistically significant ligand-receptor interaction pairs between endothelial subpopulations (EA1, EA2, and EC) and all other cell types in control, MCT, and SuHx lungs. Wilcoxon rank-sum test: $*P < 0.05$ and $**P < 0.01$. (C) Dot plots showing the log₂ mean expression of select ligand-receptor pairs in control (C), MCT (M), and SuHx (S) between select cell types and EA1 or EA2. Size of dots are proportional to strength of p-value. aMΦ = alveolar macrophages; cMono = classical monocytes; cDC = conventional dendritic cells; Fb = fibroblasts; ILC2 = type 2 innate lymphoid cells; iMΦ = interstitial macrophages; meso = mesothelial, ncMono = non-classical monocytes; NK1 = NK cells 1; NK2 = NK cells 2; PMN = neutrophils; pDC = plasmacytoid dendritic; pMΦ = proliferating macrophages; SMC = smooth muscle cells; T_prolif = proliferating T cells; T_reg = regulatory T cells.

Figure 3: Trajectory analysis reveals that EA2 has a stem/progenitor cell phenotype. (A) UMAP plots showing clustering of endothelial cells colored by CytoTRACE score (left) and cell type (right). Higher CytoTRACE scores represent a less differentiated state. (B-C) Box plots comparing cell CytoTRACE scores grouped by (B) endothelial subpopulation and (C) disease condition. Wilcoxon rank-sum test: $****P < 0.0001$. (D) Heatmap showing normalized expression of top genes most positively or negatively correlated with cell CytoTRACE scores.

Each column is a cell ordered from left to right by decreasing CytoTRACE scores. Cell annotations for cell type and condition are shown at the bottom.

Figure 4: *Tm4sf1* is a marker for hematopoietic stem cells and is upregulated in the

peripheral blood of PAH patients. (A) Box plots showing the expression of *TM4SF1* in a

dataset of 7,551 human blood cell transcriptomes representing 43 cell type clusters derived from

21 healthy donors (45). (B-D) Box plots showing expression of *TM4SF1* in PBMCs of (B) 72

PAH patients compared to 41 healthy control (46) and (C) 10 SSc-associated PAH patients

compared to 10 SSc without PAH (47) and (D) in whole blood of 65 FC II, 239 FC III, and 34

FC IV PAH patients. *P* values in (B) and (C) were obtained from GEO2R and determined by

Wilcoxon rank-sum test in (D): **P* < 0.05 and ***P* < 0.01.

Figure 5: EA2 signature genes are differentially regulated in the lungs of PAH patients and

by PAH models in human endothelial cells. (A-C) Box plots showing *TM4SF1* expression in

(A) control, MCT, and SuHx lungs (*n* = 4/group), in (B) 18 PAH patient lungs vs 13 controls

(49), and in (C) averaged endothelial scRNA-seq cells from 6 PAH patient lungs vs 3 controls

(15). (D, F) Box plots showing expression of EA2 markers *PROCR* and *SULF1* in 58 PAH

patient lungs vs 25 controls (50). (E) Box plot showing expression of *PROCR* in 18 PAH patient

lungs vs 13 controls (49). (G) Box plot showing expression of *SULF1* in 12 PAH patient lungs vs

11 controls (51). (H-K) Box plots showing expression of (H) *TM4SF1*, (I) *PROCR*, (J) *SULF1*,

(K) *NOSTRIN* in 4 *BMPR2*- vs 4 control-siRNA-transfected primary human PAECs from 4

donors (52). (L-O) Box plots showing expression of (L) *TM4SF1*, (M) *PROCR*, (N) *SULF1*, (O)

NOSTRIN in 4 *PMA*-exposed vs 4 nonexposed *BMPR2*-silenced PAECs from 4 donors (52). *P*

values were obtained from GEO2R for (B, D-O), determined by DESeq2 for (A) and by Wilcoxon rank-sum test for (C): $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****FDR < 0.05$. (P-Q) Scatter plots showing correlation between DEGs in EA2 vs. EA1 (x-axis) and DEGs from either (P) *BMPR2*-silenced PAECs ($n = 8/\text{group}$) or (Q) PMA-exposed PAECs ($n = 8/\text{group}$) (52) with select top DEGs labeled by their gene names. DEGs highlighted in yellow represent human PAH-associated genes from either (black text) or both (red text) CTD and DisGeNET databases. PMA = phorbol 12-myristate 13-acetate; logFC = log fold change; FDR = false discovery rate.

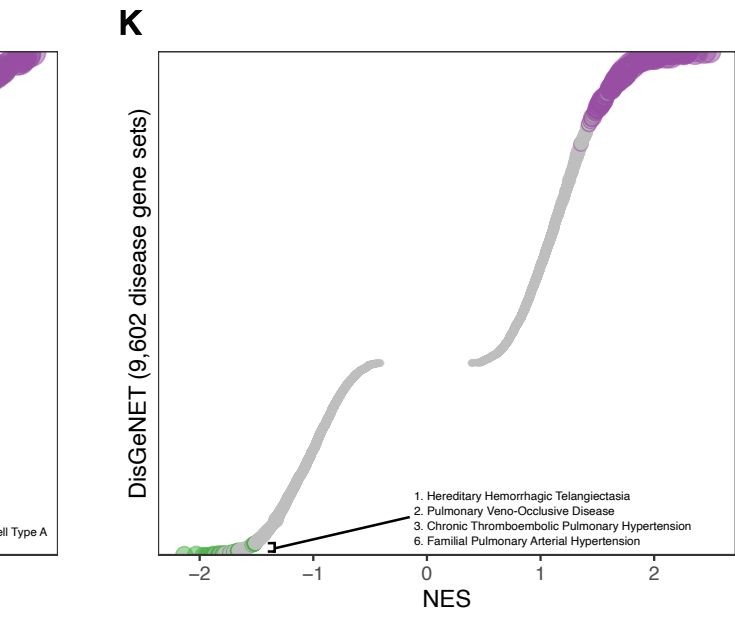
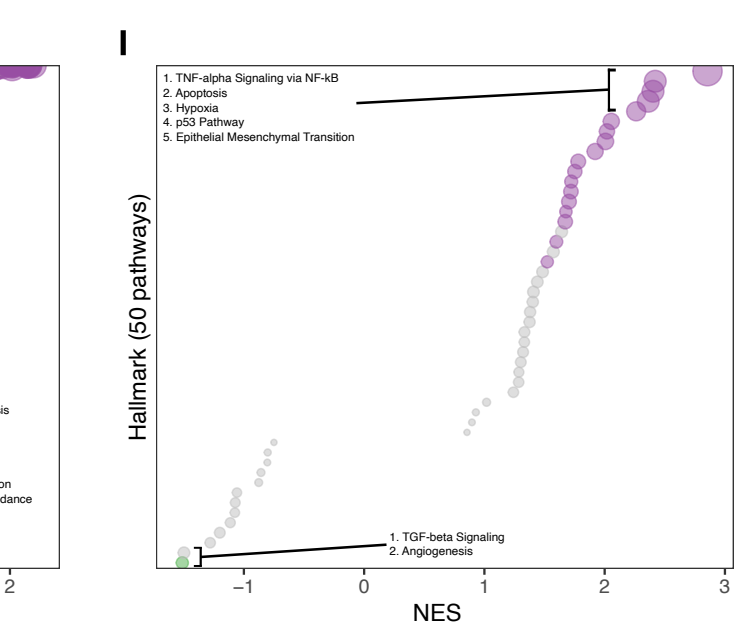


Figure 1

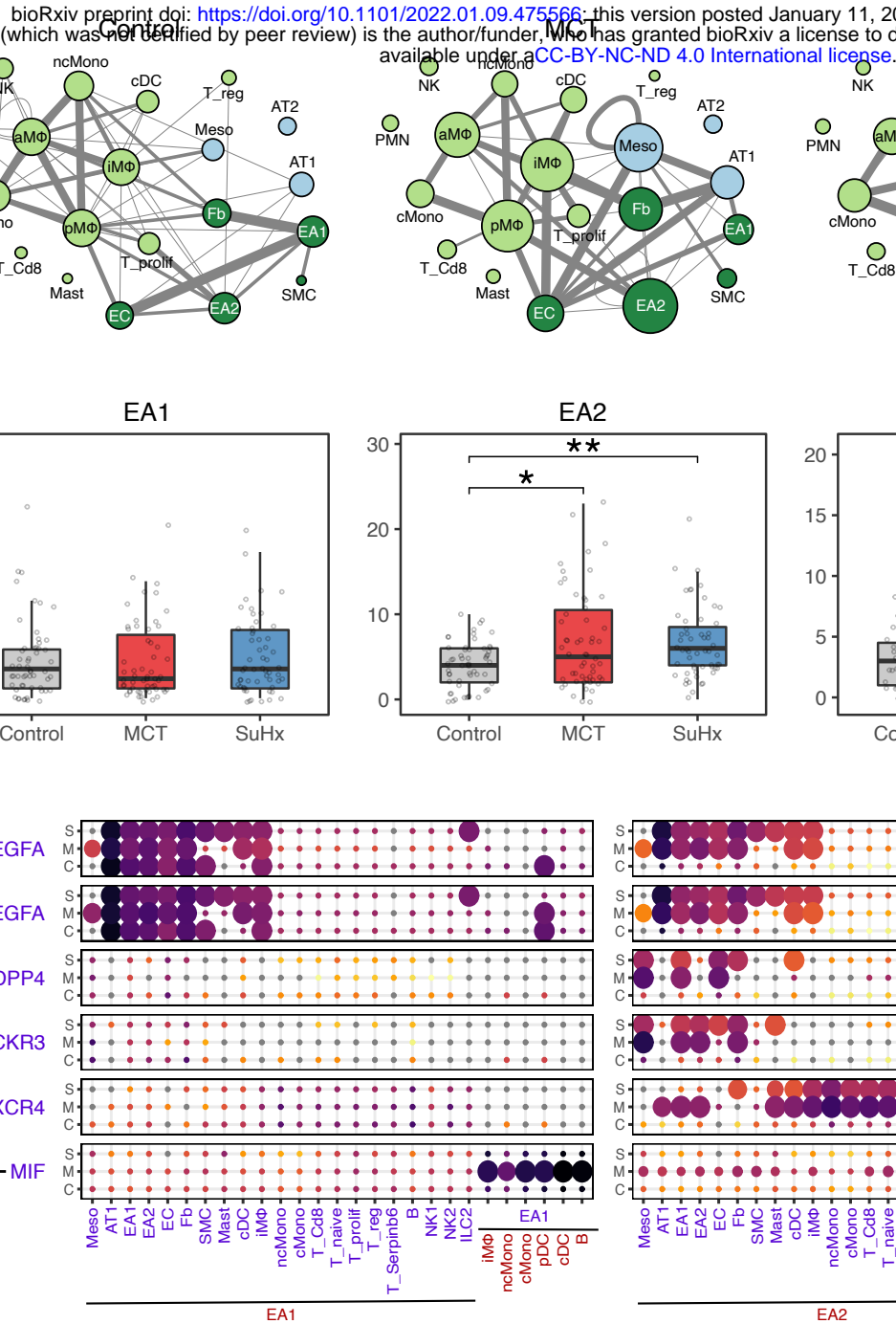


Figure 2

A

bioRxiv preprint doi: <https://doi.org/10.1101/2022.01.09.475566>; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

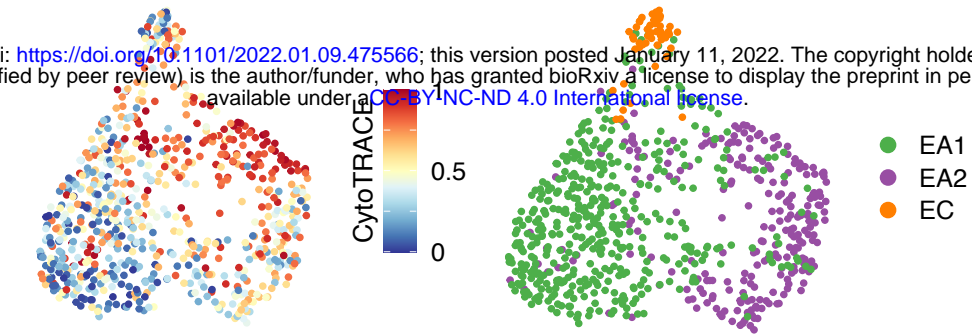
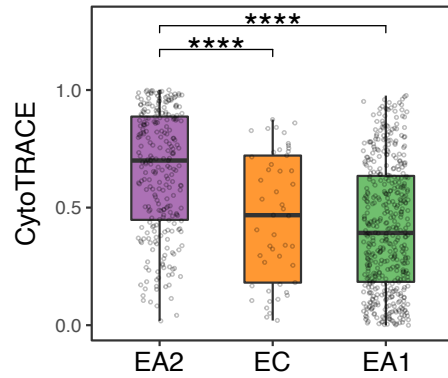
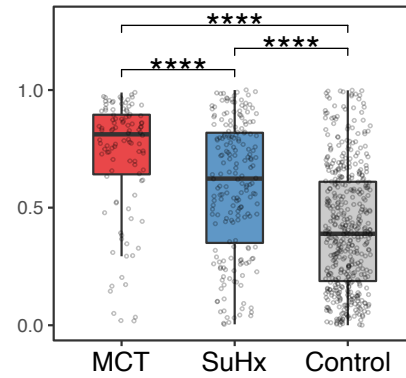
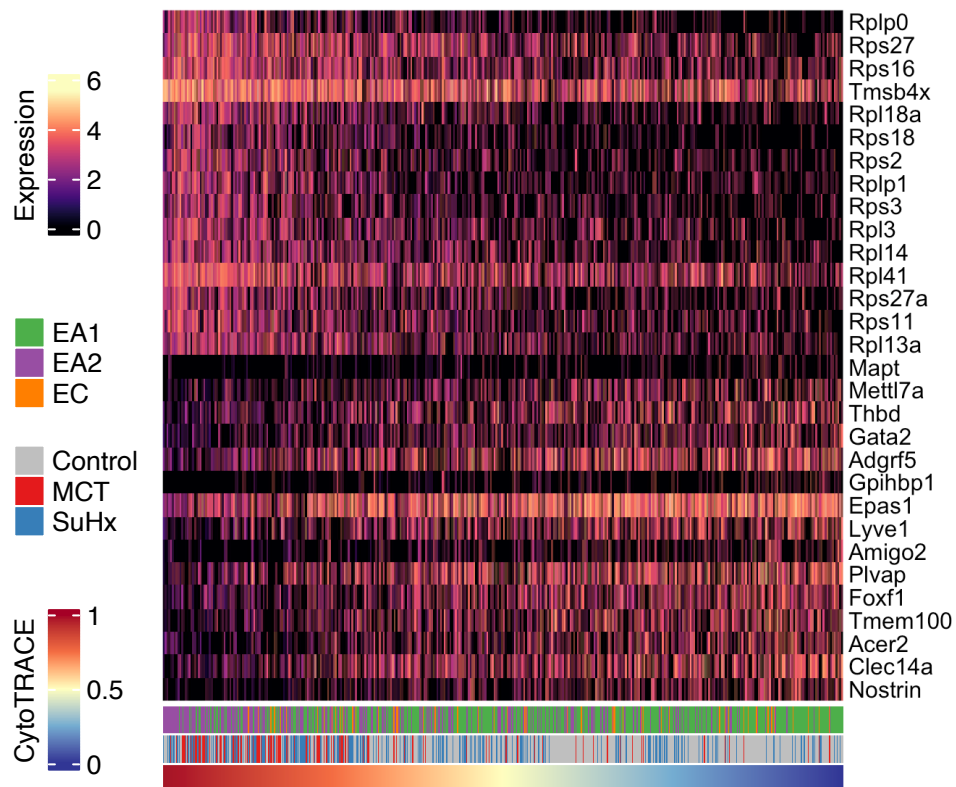
**B****C****D**

Figure 3

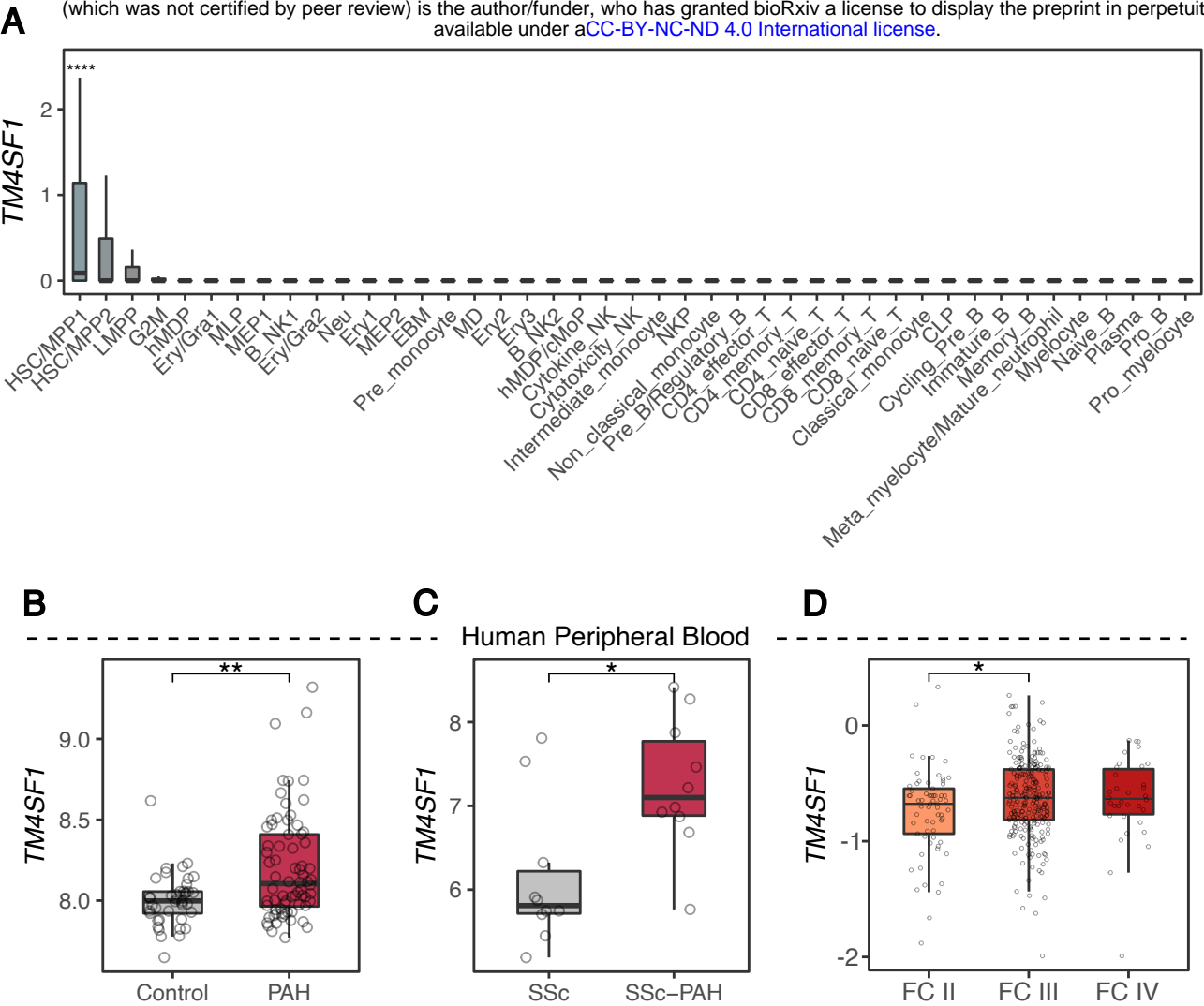
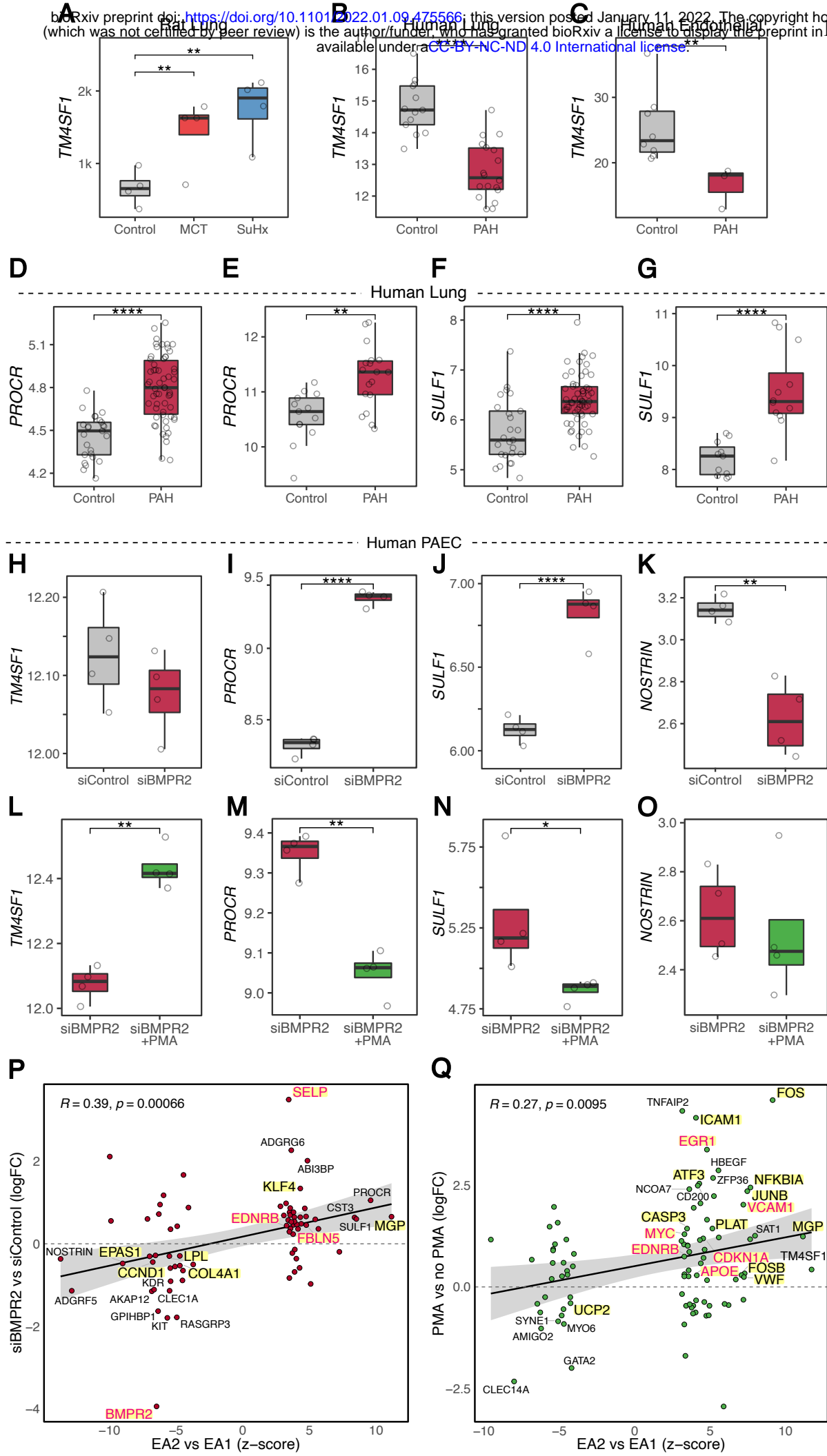


Figure 4



A Tm4sf1-Marked Subpopulation of Endothelial Stem/Progenitor Cells Identified by Lung Single-Cell Omics of Pulmonary Arterial Hypertension

Jason Hong, Brenda Wong, Caroline Huynh, Brian Tang, Soban Umar, Xia Yang, Mansoureh
Eghbali

Online Data Supplement

Supplementary Methods

Tissue processing: Lungs from control, SuHx and MCT rats were harvested after perfusion of lungs with gravity infusion of PBS via an 18g needle through the RV as previously described (1). The left lung was used for scRNA-seq and both lungs used for FACS. The tissue was minced with scissors prior to enzyme dissociation. For the first nine rats ($n = 3/\text{group}$), collagenase/dispase (Sigma, cat. no. 10269638001) at 1 mg/ml in PBS was used as previously described (2), with incubation at 37 °C for 45 min under continuous horizontal shaking (300 r.p.m.) For the second ($n = 3/\text{group}$ for scRNA-seq) and third ($n = 4/\text{group}$ for FACS) sets of rats, minced lungs were incubated with collagenase I (Worthington, cat. no. LS004194) at 1 mg/ml in DMEM + 10% FBS in a 37 °C water bath for 60 min, with trituration every 5 minutes (1–3)(3). The use of different cell dissociation protocols helps resolve potential biases in tissue dissociation and recover a more comprehensive atlas of lung cell types. After incubation, dissociated tissues were sequentially passed through 70- and 40-micron cell strainers with centrifugation at 300g for 5 min at 4 °C to pellet cells in between each step. Cells were then treated with RBC lysis buffer.

Single-cell RNA-seq barcoding and library preparation: Lung single cells from SuHx, MCT, and control rats ($n = 6/\text{group}$) were suspended at a final concentration of 100 cells/microliter in 0.01% BSA-PBS as previously described (1). Barcoded single cells, or STAMPs (single-cell transcriptomes attached to microparticles), and cDNA libraries were generated following the Drop-seq protocol from Macosko et al. (4) and version 3.1 of the online Drop-seq protocol [<http://mccarrolllab.org/download/905/>]. Briefly, single cell suspensions at 100 cells/ μl ,

EvaGreen droplet generation oil (BIO-RAD, Hercules, CA, USA), and ChemGenes barcoded microparticles (ChemGenes, Wilmington, MA, USA) were co-flowed through a FlowJEM aquapel-treated Drop-seq microfluidic device (FlowJEM, Toronto, Canada) at recommended flow speeds (oil: 15,000 $\mu\text{l/hr}$, cells: 4000 $\mu\text{l/hr}$, and beads 4000 $\mu\text{l/hr}$) to generate STAMPs. The following modifications were made to the online published protocol to obtain enough cDNA as quantified by TapeStation (Agilent) to continue the protocol: (1) the number of beads in a single PCR tube was 4000, (2) the number of PCR cycles was 4 + 11 cycles, and (3) multiple PCR tubes were pooled. The libraries were then checked on TapeStation for library quality, average size, and concentration estimation. The samples were then tagmented using the Nextera DNA Library Preparation kit (Illumina) and multiplex indices were added. After another round of PCR, the samples were checked on TapeStation for library quality check before sequencing. (A cell doublet rate of 5.6% was obtained by running the microfluidic device without the lysis buffer and counting the percentage of cell doublets through three separate runs.)

Illumina high-throughput sequencing of single-cell RNA-seq libraries: The Drop-seq library molar concentration was quantified by Qubit Fluorometric Quantitation (Thermo Fisher) and library fragment length was estimated using TapeStation as previously described (1). Sequencing was performed on an Illumina HiSeq 4000 (Illumina) instrument using the Drop-seq custom read 1B primer (GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC) (IDT, Coralville, IA, USA). Paired end reads were 100bp in length. Read 1 consists of the 12 bp cell barcode, followed by the 8 bp UMI, and the last 80 bp on the read are not used. Read 2 contains the single cell transcripts.

Single-cell RNA-seq data pre-processing and quality control: Demultiplexed fastq files

generated from Drop-seq were processed to digital expression gene matrices using Drop-seq tools version 1.13 (<https://github.com/broadinstitute/Drop-seq>) as previously described (1). The workflow is available on github (<https://github.com/darneson/dropSeqPipeDropEST>). Briefly, fastq files were converted to BAM format and cell and molecular barcodes were tagged. Reads corresponding to low quality barcodes were removed and any occurrence of the SMART adapter sequence or polyA tails found in the reads was trimmed. These cleaned reads were converted back to fastq format to be aligned to the rat reference genome Rnor 6.0 using STAR-2.5.0c. After the reads were aligned, the reads which overlapped with exons, introns, and intergenic regions were tagged using a RefFlat annotation file of Rnor 6.0. The count values for each cell were normalized by the total number of UMIs in that cell and then multiplied by 10,000 and log transformed. Single cells were identified from background ambient mRNA using thresholds of at least 200 genes, 400 transcripts and no more than 50% mitochondrial fraction.

Identification of cell clusters: The Seurat R package version 3.0.0.9000

(<https://github.com/satijalab/seurat>) was used to project all sequenced cells onto two dimensions using Uniform Manifold Approximation and Projection (UMAP) (5) and Louvain clustering (6) was used to assign clusters as previously described (1). The optimal number of principal components used for UMAP dimensionality reduction and Louvain clustering was determined using the Jackstraw permutation approach. The density used to assign clusters was identified using a parameter grid search.

Identification of marker genes and cell types of individual cell clusters: We defined cell cluster specific marker genes from our Drop-seq dataset using the FindConservedMarkers function in Seurat across all the samples as previously described (1). Briefly, a Wilcoxon rank-sum test is run within each sample and a meta p-value across all samples is computed to assess the significance of each gene as a marker for a cluster. Within each sample, the cells are split into two groups: single cells from the cell type of interest and all other single cells. To be considered in the analysis, the gene had to be expressed in at least 10% of the single cells from one of the groups and there had to be at least a 0.25 log fold change in gene expression between the groups. This process was conducted within each sample separately, and then a meta p-value was assessed from the p-values across all samples. Multiple testing was corrected using the Benjamini-Hochberg method on the meta p-values and genes with an FDR < 0.05 were defined as cell type-specific marker genes. Cell types were identified based on cell type-specific marker genes from ImmGen (7), LungMAP (8), and Mouse Cell Atlas (3). Additional integrative methods, canonical correlation analysis and mutual nearest neighbors (9, 10), were also used to confirm and optimize cell clustering.

Bulk RNA-seq: The UCLA Technology Center for Genomics & Bioinformatics (TCGB) generated libraries from total RNA and sequenced the libraries as paired-end 50 base pair reads using Hiseq 3000 (Illumina). Reads were aligned to Rnor 6.0 genome using HISAT2 version 2.1.0 and transcripts were assembled and quantified using StringTie version v1.3.3b (11).

Identification of differentially expressed genes (DEGs) and pathways: DEG analysis was performed using MAST (12) and a false discovery rate (FDR) < 0.05 was considered statistically

significant. Pathway enrichment analysis was performed using Gene Ontology (GO) (13), Hallmark pathways from Molecular Signature Database (14), Human Biomolecular Atlas Program (HuBMAP) (15), and DisGeNET (16) in R package fgsea v1.18.0 with genes ordered by MAST-derived Z scores. Human PAH-associated gene sets were obtained from DisGeNET (16) and Comparative Toxicogenomics Database (17) using the Harmonizome portal (18). For bulk RNA-seq data, DEG analysis was performed using DESeq2 (19) and a false discovery rate (FDR) < 0.05 was considered statistically significant. Enrichment of Hallmark pathways for overlapping genes between DEGs in EA2 (vs. EA1) and DEGs in human PAECs after either *BMPR2* knockdown or PMA stimulation were determined by hypergeometric test as implemented by Molecular Signatures Database at <https://www.gsea-msigdb.org/gsea/msigdb/annotate.jsp>.

Independent datasets for validation: Select genes were queried in independent bulk and single-cell transcriptomic datasets of human lung and blood samples from Gene Expression Omnibus (GEO) or hosted on lab web servers. For the whole blood RNA-seq dataset (20), *Tm4sf1* expression was queried by functional class at <https://sheffield-university.shinyapps.io/ipah-rnaseq-app/> to generate a PDF file showing individual datapoints along a log₁₀(TPM) expression y-axis. The PDF was then imported into Inkscape version 1.1.1 after which the expression y-axis was proportioned to Inkscape y coordinates to determine expression values of individual data points.

Table E1: Genes differentially expressed in endothelial arterial type 2 (EA2) compared to endothelial arterial type 1 (EA1). Z-scores were derived from MAST where scores above 0

denote upregulated genes and scores below 0 represent downregulated genes. Genes with FDR < 0.05 are listed.

<i>Gene</i>	<i>z-score</i>	<i>p-value</i>	<i>FDR</i>
Tm4sf1	11.7282975	6.05E-36	1.22E-31
Mgp	11.154162	2.21E-32	2.22E-28
Adgrf5	-12.826645	6.64E-31	4.46E-27
Nostrin	-13.665406	7.08E-30	3.57E-26
Procr	9.58656079	1.87E-28	7.55E-25
Sulf1	8.51380349	3.14E-24	1.06E-20
Epas1	-9.0082233	2.03E-22	5.83E-19
AY172581.24	-11.159437	5.77E-22	1.45E-18
Acer2	-10.590517	3.51E-21	7.86E-18
Plvap	-9.5373122	9.83E-20	1.98E-16
Fos	9.1454617	1.57E-18	2.87E-15
Sema3c	-9.9730334	5.09E-18	8.54E-15
Slc29a1	-9.8851676	9.81E-18	1.52E-14
Vcam1	7.20208733	3.12E-17	4.49E-14
Cst3	8.37949102	9.15E-16	1.23E-12
Nfkbiz	7.65297342	5.11E-15	6.43E-12
Fosb	7.28791331	4.90E-14	5.80E-11
Clec14a	-7.995728	5.98E-14	6.69E-11
Foxf1	-7.1485609	7.50E-14	7.95E-11
Sat1	7.9788583	1.11E-13	1.11E-10
Myof	6.22533668	2.66E-13	2.55E-10
Eln	6.70121875	3.01E-13	2.76E-10
Nfkbia	7.67195981	4.46E-13	3.91E-10
Nuak1	5.89726286	1.10E-12	9.25E-10
RT1-CE4	-6.4308983	1.26E-12	1.01E-09
Vwf	7.24243386	4.46E-12	3.46E-09
Junb	7.46543679	5.04E-12	3.76E-09
Jam2	7.08069107	5.96E-12	4.29E-09
Anxa3	7.20717246	6.78E-12	4.71E-09
Fbln2	5.43482631	7.51E-12	4.88E-09
Ppap2a	6.52867137	7.44E-12	4.88E-09
Igfbp7	-5.9403516	1.11E-11	6.99E-09
Cdh5	-6.9868196	1.56E-11	9.55E-09
Wif1	-7.5933096	4.28E-11	2.54E-08

Ch25h	5.25158568	4.88E-11	2.80E-08
Sptbn1	-5.4349312	5.01E-11	2.80E-08
Cmip	6.11692901	1.13E-10	6.06E-08
Gja5	5.77870543	1.14E-10	6.06E-08
Akap12	-6.8258704	1.96E-10	1.01E-07
Gpx2	-7.1976013	3.05E-10	1.54E-07
Ccdc68	-6.2947513	4.79E-10	2.35E-07
Sdpr	-6.2055416	5.71E-10	2.74E-07
BC028528	-6.971917	1.29E-09	6.03E-07
S100a10	5.96015635	1.32E-09	6.06E-07
Mt2A	5.78827432	1.50E-09	6.72E-07
Ccnd1	-6.7503965	2.51E-09	1.10E-06
Bmpr2	-6.4673306	2.84E-09	1.22E-06
S100a6	5.98280232	3.91E-09	1.64E-06
Hbegf	5.55236382	5.02E-09	2.06E-06
Slfn5	-5.6439159	6.94E-09	2.74E-06
Cyyr1	-6.455843	7.41E-09	2.87E-06
Kdr	-6.6538338	8.32E-09	3.16E-06
Cdc42ep1	-6.5459263	1.15E-08	4.29E-06
Tmem100	-6.2620781	1.56E-08	5.73E-06
LOC500300	5.09146195	1.89E-08	6.79E-06
Emp1	5.57456186	2.13E-08	7.54E-06
Fbln5	4.76193656	2.43E-08	8.45E-06
Amigo2	-6.2060006	2.73E-08	9.32E-06
Pnrc1	5.67142367	2.91E-08	9.76E-06
Fn1	5.01422746	3.03E-08	9.96E-06
Ltbp4	5.09919652	3.06E-08	9.96E-06
Tinag11	4.64897855	3.55E-08	1.13E-05
Aplnr	-6.0469545	5.23E-08	1.64E-05
Clic5	-6.1772171	5.38E-08	1.67E-05
Fgl2	4.91151918	5.92E-08	1.81E-05
Gpihbp1	-6.350405	6.02E-08	1.81E-05
Mt-co1	-5.6265234	6.35E-08	1.88E-05
Btg2	4.79316297	7.48E-08	2.18E-05
Dpp4	-6.0030288	8.71E-08	2.51E-05
Cdh13	4.3117471	1.48E-07	4.19E-05
Lrrfip1	4.95690365	2.20E-07	6.15E-05
Kit	-5.6574857	2.98E-07	8.23E-05
Crip1	5.10045534	3.04E-07	8.27E-05
Enc1	4.38085558	4.37E-07	0.00011573

Zfp36	5.19611625	4.50E-07	0.00011759
Plac8	4.30214679	6.85E-07	0.00017701
Dhh	4.14141215	8.80E-07	0.0002245
Glp1r	-5.7057387	9.24E-07	0.00023274
Anxa1	4.74537374	9.36E-07	0.00023289
Abi3bp	4.84878305	9.64E-07	0.00023677
Lpl	-5.5051869	1.01E-06	0.00024447
Egr1	4.78859223	1.14E-06	0.00027224
Arap2	-5.2126666	1.16E-06	0.00027417
Mcam	-5.4073346	1.19E-06	0.00027955
Ppap2b	4.83072947	1.37E-06	0.00031697
Plat	5.15075642	1.41E-06	0.00031884
Sh2d3c	-5.4130644	1.40E-06	0.00031884
Il33	4.3601416	1.75E-06	0.00039224
Clec1a	-5.4906752	1.90E-06	0.00042033
Rdx	-5.2844395	2.85E-06	0.00062338
Timp3	4.01569539	2.92E-06	0.0006331
Tspan8	-5.4178543	3.05E-06	0.00065368
Akr1cl	-5.0962927	4.14E-06	0.00087839
Gsn	4.26055923	4.96E-06	0.00104101
Klf4	4.31311427	5.56E-06	0.0011542
Mt-atp6	-4.8038251	6.21E-06	0.00126414
Myo10	-5.3519987	6.16E-06	0.00126414
Cd200	4.1820317	6.63E-06	0.0013363
Pde4d	3.8653914	7.27E-06	0.00144944
Car4	-5.2089379	7.83E-06	0.00153035
Bank1	-5.1859877	8.02E-06	0.00155281
Dkk2	3.50838755	8.25E-06	0.00158205
Ptprj	3.87647789	8.77E-06	0.00166621
Dst	4.6147836	9.23E-06	0.0017383
Hspa1b	3.2883167	1.02E-05	0.00186545
Selp	3.44450751	1.00E-05	0.00186545
Cdkn1a	4.54085517	1.05E-05	0.00189053
Scn7a	-5.2106626	1.05E-05	0.00189053
Fbn1	3.88441073	1.19E-05	0.00211428
Syne1	-5.078818	1.36E-05	0.00240271
Ubc	4.79655512	1.37E-05	0.00240535
Fxyd5	4.05228782	1.41E-05	0.0024544
Ppp1r15a	4.56298937	1.48E-05	0.0025468
S100a4	3.48497117	1.53E-05	0.00260728

Tmcc2	-5.1039366	1.56E-05	0.00263931
Rasgrp3	-4.9474318	1.95E-05	0.00326537
Ppl	3.57331694	1.98E-05	0.00330362
Slco2a1	-4.9175986	2.09E-05	0.00344627
Tnfrsf12a	3.63656147	2.23E-05	0.00364825
Dap	3.80343843	2.69E-05	0.00434237
Synm	-4.8403677	2.77E-05	0.00443595
Slfn3	3.35266081	2.82E-05	0.00447203
Mpp3	-4.3047837	3.01E-05	0.00474021
Adgrg1	3.56975684	3.06E-05	0.00474208
Csrp2	3.26676386	3.05E-05	0.00474208
Ccnl1	4.63455097	3.26E-05	0.00501042
Tspan7	-3.5451301	3.29E-05	0.00501723
Atf3	4.28212729	3.32E-05	0.00502725
Col4a1	-4.717117	3.79E-05	0.00570083
Adgrg6	3.63757923	4.39E-05	0.00651689
Slc6a2	3.56632504	4.40E-05	0.00651689
Hsp90ab1	3.09601681	5.17E-05	0.00754435
Tppp3	-4.8203248	5.37E-05	0.00778478
Rgs12	-4.7219638	5.49E-05	0.00784132
Lmo7	3.59166163	5.83E-05	0.00827269
Apoe	4.11562249	6.35E-05	0.00893808
Icam1	4.0590773	6.91E-05	0.00966413
Efnb2	-4.5338312	7.19E-05	0.00998933
Myo6	-4.6937828	7.75E-05	0.01069627
Nr4a2	3.91247943	8.57E-05	0.01166932
Gpx1	4.06984325	8.84E-05	0.01195533
Casp3	3.47095551	9.23E-05	0.01231371
PVR	3.6205546	9.56E-05	0.01267115
Cxcl16	3.63320719	9.83E-05	0.0128551
Ehbp111	3.29633814	0.00010227	0.01329096
Swap70	3.49416398	0.00010831	0.01398463
Ier3	4.32273662	0.00010958	0.01405892
Mecom	3.49324857	0.00012729	0.01612528
Efemp1	3.245051	0.0001298	0.01634143
Gmfg	-4.5234732	0.00013133	0.01643101
Srpx	3.55647244	0.00013528	0.01682045
Afap112	3.13667223	0.00014058	0.0171947
Kcne3	3.51737069	0.00014023	0.0171947
Mapt	-4.4574214	0.00014085	0.0171947

Hspb1	4.03910106	0.00014295	0.01734617
Gata2	-4.1868194	0.00015148	0.01827074
Abcb1a	-4.5341441	0.00015557	0.01865235
Nrp1	-4.2706811	0.00016067	0.01914989
Hopx	-4.4476375	0.00016844	0.01995782
Cpe	3.22396019	0.00017114	0.02016005
Arhgdia	3.71366798	0.00017279	0.02023554
Anks1a	-4.4839102	0.00019942	0.02321948
Bcl2a1	3.43626948	0.00021364	0.02473235
Rab27a	3.57883867	0.00022212	0.02556701
Myc	3.29920723	0.00022753	0.02604069
Ptprc	3.47196475	0.00023254	0.02646363
Gadd45g	3.84818833	0.000274	0.03094999
RT1-CE5	-4.3401776	0.00027504	0.03094999
Sept4	-4.0734169	0.00028206	0.03138974
Tes	3.84740102	0.00028716	0.03178221
Slc1a5	3.10178179	0.00029091	0.03202056
Sorbs2	3.36179884	0.00030072	0.0329207
Ip6k2	3.28871465	0.00030626	0.03334569
Tyrobp	3.33318356	0.0003099	0.03356077
Cd63	3.90196167	0.00031357	0.03359727
Olfml2a	-4.2979733	0.0003126	0.03359727
RGD1563861	3.72674639	0.00031942	0.03393319
Tnfaip2	3.15657515	0.00032008	0.03393319
Add3	-4.2659282	0.00032318	0.03408239
Lmna	3.7729458	0.00033897	0.03556207
Ucp2	-4.2551658	0.00034165	0.03565726
Ifi27	-3.717774	0.00035004	0.03615856
Slc43a3	-4.0995841	0.00034853	0.03615856
Prss23	3.64602682	0.00035206	0.03618143
Ncoa7	3.62114399	0.00037068	0.03770985
Mt1	3.49519032	0.00038524	0.03899454
Itga1	-4.2381472	0.00038985	0.03926386
RT1-A1	-3.901491	0.00039332	0.03941637
Cks1b	2.82236042	0.00042166	0.04183971
Htra1	3.40585123	0.00043903	0.04334982
Ednrb	3.28617831	0.00045072	0.04428704
Rac2	3.08059323	0.00046133	0.04510957
Igf2	3.22187733	0.00047679	0.04639605
Ctsl	-3.9958321	0.0004795	0.04643591

Prx	-3.9251525	0.00048296	0.04654663
Stk4	-4.030701	0.00049884	0.04784806

References

1. Hong J, Arneson D, Umar S, Ruffenach G, Cunningham CM, Ahn IS, Diamante G, Bhetraratana M, Park JF, Said E, Huynh C, Le T, Medzikovic L, Humbert M, Soubrier F, Montani D, Girerd B, Trégouët D-A, Channick R, Saggar R, Eghbali M, Yang X. Single-Cell Study of Two Rat Models of Pulmonary Arterial Hypertension Reveals Connections to Human Pathobiology and Drug Repositioning. *Am J Respir Crit Care Med* 2021;203:1006–1022.
2. Alphonse RS, Vadivel A, Zhong S, Zong S, McConaghy S, Ohls R, Yoder MC, Thébaud B. The isolation and culture of endothelial colony-forming cells from human and rat lungs. *Nat Protoc* 2015;10:1697–1708.
3. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, Huang D, Xu Y, Huang W, Jiang M, Jiang X, Mao J, Chen Y, Lu C, Xie J, Fang Q, Wang Y, Yue R, Li T, Huang H, Orkin SH, Yuan G-C, Chen M, Guo G. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* 2018;172:1091-1107.e17.
4. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 2015;161:1202–1214.
5. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:180203426 [cs, stat]* 2018;at <<http://arxiv.org/abs/1802.03426>>.

6. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech* 2008;2008:P10008.
7. Heng TSP, Painter MW, Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* 2008;9:1091–1094.
8. Ardini-Poleske ME, Clark RF, Ansong C, Carson JP, Corley RA, Deutsch GH, Hagood JS, Kaminski N, Mariani TJ, Potter SS, Pryhuber GS, Warburton D, Whitsett JA, Palmer SM, Ambalavanan N, LungMAP Consortium. LungMAP: The Molecular Atlas of Lung Development Program. *Am J Physiol Lung Cell Mol Physiol* 2017;313:L733–L740.
9. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. *Cell* 2019;177:1888–1902.e21.
10. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology* 2018;36:421–427.
11. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 2016;11:1650–1667.
12. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, Linsley PS, Gottardo R. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 2015;16:.

13. The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, Basu S, Chisholm RL, Dodson RJ, Hartline E, Fey P, Thomas PD, Albou L-P, Ebert D, Kesling MJ, Mi H, Muruganujan A, Huang X, Mushayahama T, LaBonte SA, Siegele DA, Antonazzo G, Attrill H, Brown NH, Garapati P, Marygold SJ, Trovisco V, dos Santos G, *et al.* The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* 2021;49:D325–D334.
14. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–425.
15. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* 2019;574:187–192.
16. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 2017;45:D833–D839.
17. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegiers J, Wiegiers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res* 2019;47:D948–D954.
18. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma’ayan A. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* 2016;2016:.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 2014;15:550.

20. Kariotis S, Jammeh E, Swietlik EM, Pickworth JA, Rhodes CJ, Otero P, Wharton J, Iremonger J, Dunning MJ, Pandya D, Mascarenhas TS, Errington N, Thompson AAR, Romanoski CE, Rischard F, Garcia JGN, Yuan JX-J, An T-HS, Desai AA, Coghlan G, Lordan J, Corris PA, Howard LS, Condliffe R, Kiely DG, Church C, Pepke-Zaba J, Toshner M, Wort S, *et al.* Biological heterogeneity in idiopathic pulmonary arterial hypertension identified through unsupervised transcriptomic profiling of whole blood. *Nat Commun* 2021;12:7104.

Supplementary Figure Legends

Figure E1: Endothelial expression of arterial marker *Efnb2* and capillary marker *Car4*.

Violin plots showing expression of (A) *Efnb2* and (B) *Car4* in rat lung scRNA-seq grouped by endothelial subpopulations.

Figure E2: Violin plot showing expression of *Tm4sf1* across all rat lung scRNA-seq cell types.

Figure E3: Ligand-receptor interactions in EC. Dot plot showing the log2 mean expression of select ligand-receptor pairs in control (C), MCT (M), and SuHx (S) between select cell types and EC. Size of dots are proportional to strength of p-value.

Figure E4: Biological processes enriched in genes correlated with a less differentiated state of endothelial cells. Dots plot showing GSEA of the CytoTRACE gene signature (i.e. genes

correlated with CytoTRACE scores) using GO. The x-axis represents normalized enrichment scores (NES). Dots with $FDR < 0.05$ are colored in red and represent gene sets significantly enriched in genes correlated with differentiation state. The y-axis represents gene sets ordered by their NES. Select gene sets are labeled and numbered by their ordering as top gene sets enriched in genes correlated with a less differentiated state. Dots larger in size represent higher $-\log_{10}(FDR)$ values.

Figure E5: EA2 signature genes are differentially regulated by *BMPR2* knockdown and by PMA exposure in human endothelial cells. (A-D) Box plots showing expression of (A) *TM4SF1*, (B) *PROCR*, (C) *SULF1*, (D) *NOSTRIN* in 4 *BMPR2*- vs 4 control-siRNA-transfected primary human PAECs exposed to PMA from 4 donors (52). (E-H) Box plots showing expression of (E) *TM4SF1*, (F) *PROCR*, (G) *SULF1*, (H) *NOSTRIN* in 4 PMA-exposed vs 4 nonexposed siControl PAECs from 4 donors (52). PMA = phorbol 12-myristate 13-acetate.

Figure E5: Overlapping DEGs between EA2 and *in vitro* PAH models are enriched for known PAH pathways. Bar plots showing enrichment of Hallmark pathways for overlapping genes between DEGs in EA2 (vs. EA1) and DEGs in human PAECs after either (A) *BMPR2* knockdown or (B) PMA stimulation. Enrichment was determined by hypergeometric test. FDR = false discovery rate.

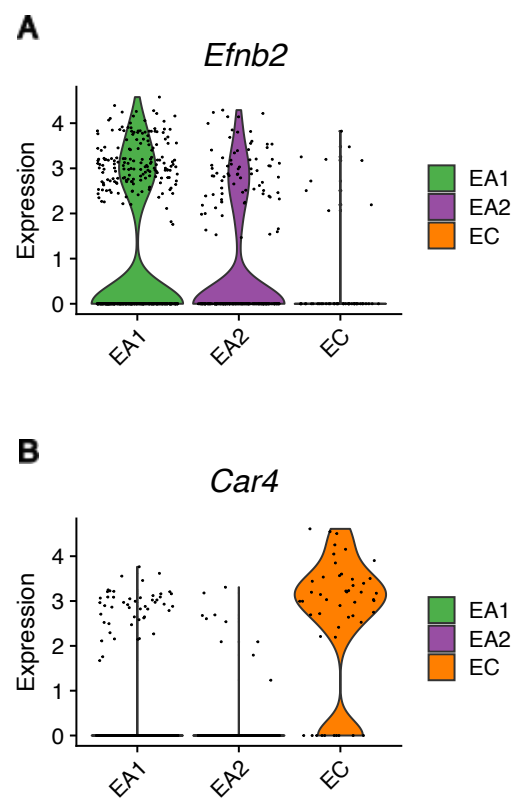


Figure E1

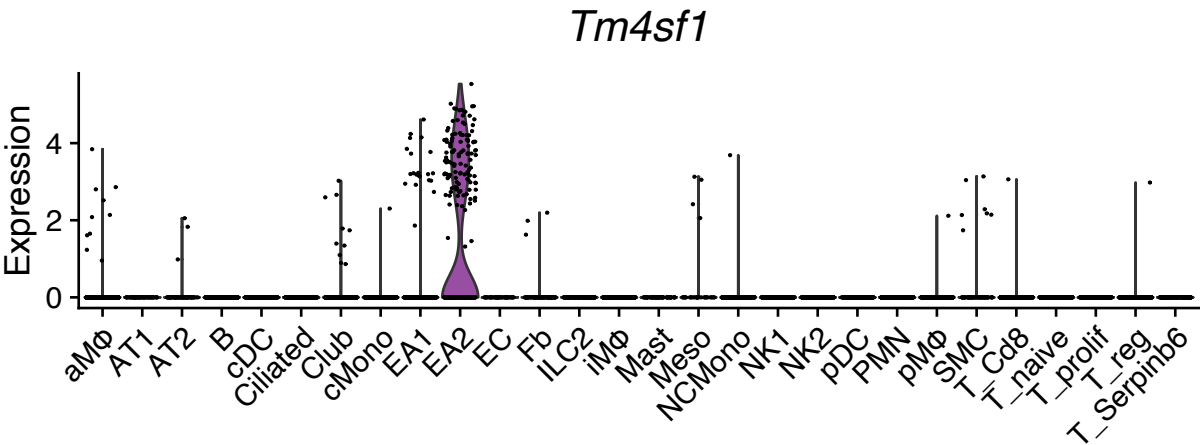


Figure E2

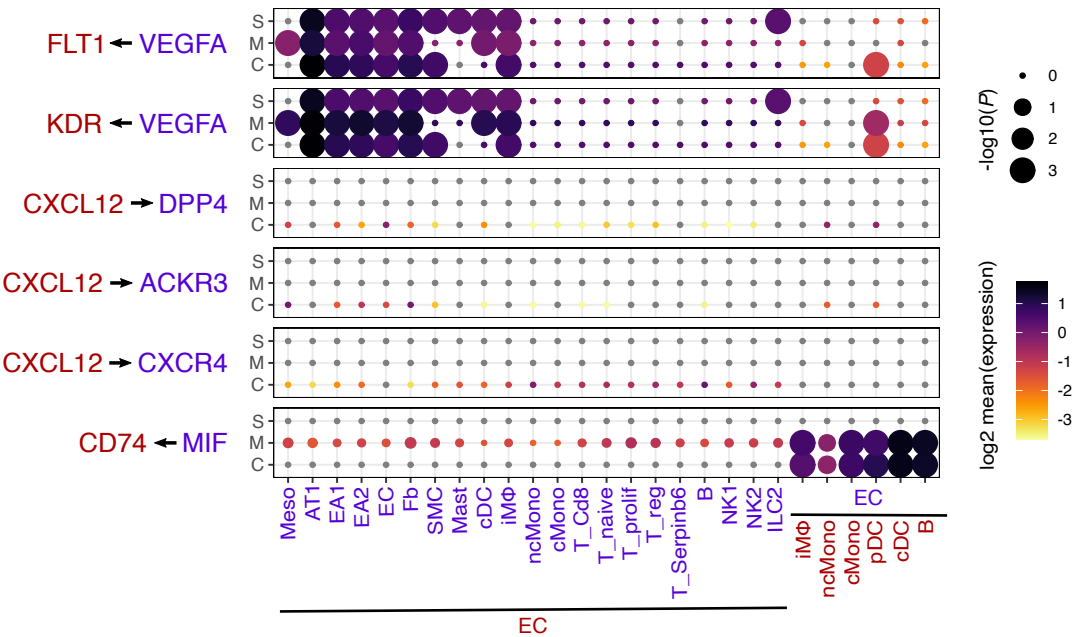


Figure E3

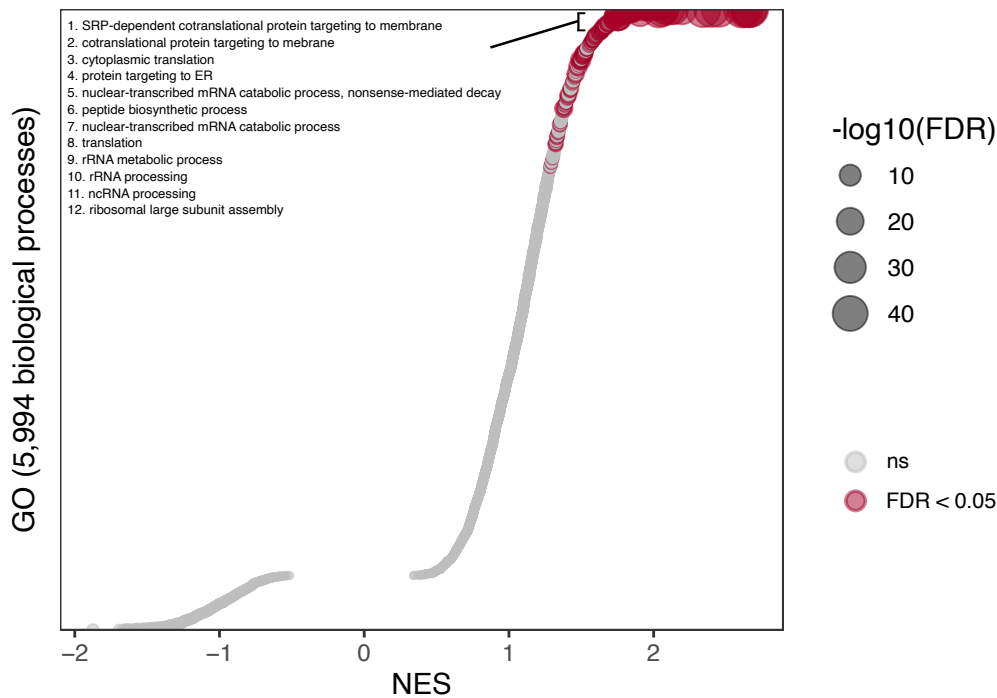


Figure E4

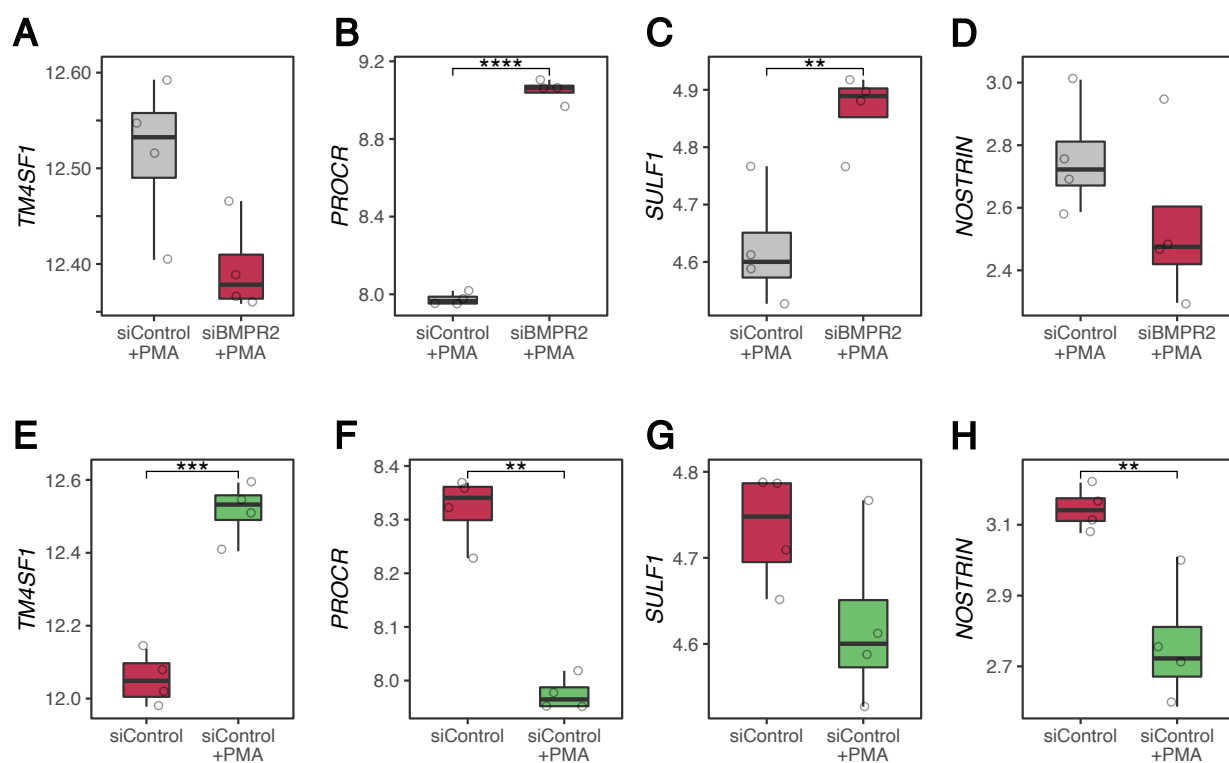
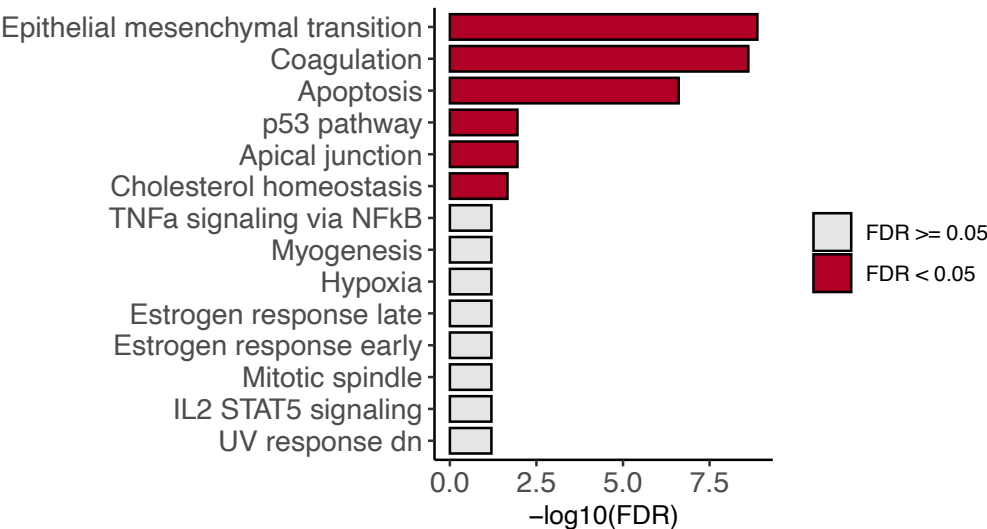


Figure E5

A



B

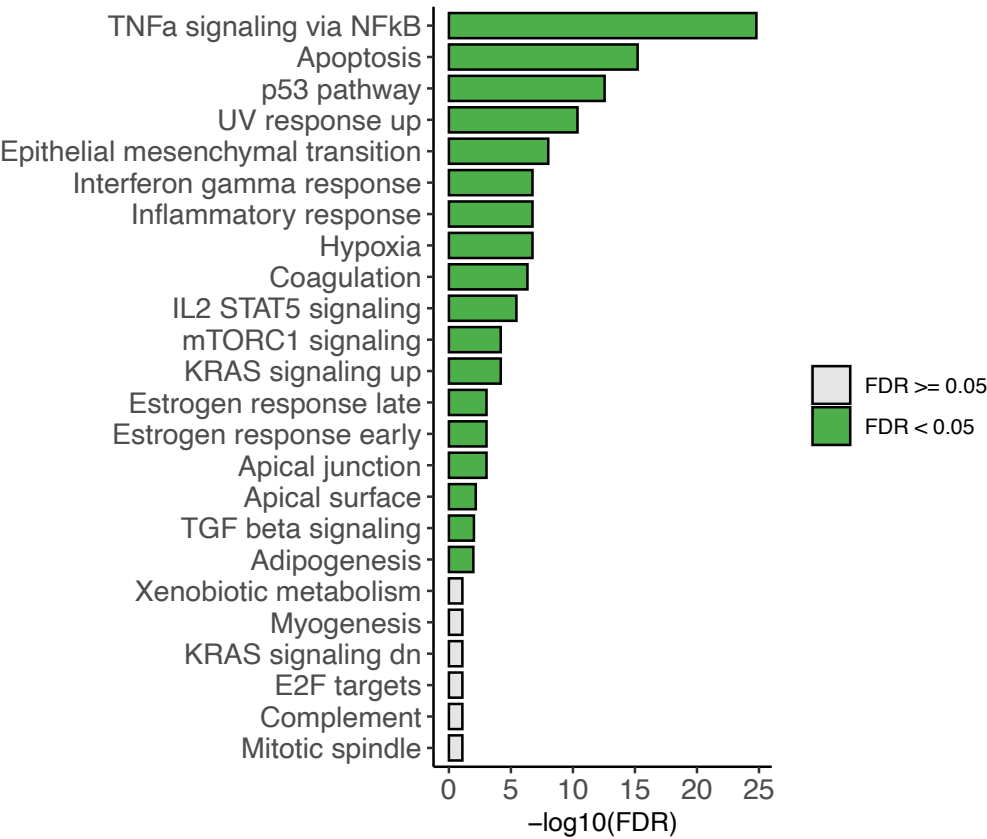


Figure E6