Emergence of disease specific endothelial and stromal cell populations involved in arterial remodeling during development of pulmonary arterial hypertension

Nicholas D Cober[1,2], Emma McCourt[1], Rafael Soares Godoy[1], Yupu Deng[1], Ken Schlosser[1], Anu Situ[1], David P Cook[2], Sarah-Eve Lemay[3], Timothy Klouda[4], Ke Yuan[4], Sébastien Bonnet[3], and Duncan J Stewart[1,2,*]

[1]Ottawa Hospital Research Institute, Sinclair Centre for Regenerative Medicine Program, Ottawa, ON, Canada
[2]University of Ottawa, Faculty of Medicine, Department of Cellular and Molecular Medicine, Ottawa, ON, Canada
[3]Institut Universitaire de Cardiologie et de Pneumologie de Quebec. Research Center Laval University, Quebec City, QC, Canada
[4]Division of Pulmonary Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

ORCIDs:

NDC: 0000-0001-8061-806X
DPC: 0000-0001-7639-6724
DJS: 0000-0002-9113-8691

Short title: Single cell atlas of PH progression

Total Word Count: 6696

Abstract word count: 281

Total figure count: 8

***Corresponding author:** Duncan J Stewart, djstewart@ohri.ca.
Executive Vice-President Research, The Ottawa Hospital; CEO & Scientific Director, Ottawa Hospital Research Institute; The Evelyne & Rowell Laishley Chair and Professor, Department of Medicine, University of Ottawa
501 Smyth Road, Box/C.P. 511, Ottawa, ON, K1H 8L6
613-739-6686

Abstract

Pulmonary arterial hypertension (PAH) is a severe and lethal pulmonary vascular disease characterized by arteriolar pruning and occlusive vascular remodeling leading to increased pulmonary vascular resistance and eventually right heart failure. While endothelial cell (EC) injury and apoptosis are known triggers for this disease, the mechanisms by which they lead to complex arterial remodeling remain obscure. We employed multiplexed single-cell RNA sequencing (scRNA-seq) at multiple timepoints during the onset and progression of disease in the SU/CH model of PAH to identify mechanisms involved in the development of occlusive arterial lesions. There was significant loss of arterial volume as early as 1-week by microCT, preceding any evidence of occlusive arteriopathy, consistent with early arteriolar dropout. Maximal arterial pruning was seen by 5 to 8 weeks, with signs of progressive occlusive remodeling. Analysis of the scRNA-seq data resolved 44 lung cell populations, with widespread transcriptional changes at 1 week affecting endothelial, stromal, and immune cell populations. Notably, this included emergence of a relatively dedifferentiated general capillary (gCap) EC population (dCap ECs) that expresses *Cd74* and appears to undergo endothelial-mesenchymal transition by RNA velocity analysis. However, at late timepoints (5 and 8 weeks), only the activated arterial EC (aAEC) population exhibited persistent differential gene expression characterized by a dysfunctional growth regulated state. This included high expression of *Tm4sf1*, implicated in cancer cell growth, which was also found in a smooth muscle (SM)-like pericyte population. Both aAECs and SM-like pericytes were found to be spatially localized to regions of occlusive arterial remodeling in lungs of SU/CH rats and PAH patient samples.

Together these findings implicate disease specific vascular cells in PAH progression and suggest that TM4SF1 may be a novel therapeutic target for arterial remodeling.

Keywords: Pulmonary arterial hypertension; Endothelial cells; Pericytes; Vascular remodeling

Introduction

Pulmonary arterial hypertension (PAH) is a severe and lethal disease defined hemodynamically by a mean pulmonary artery pressure > 20mmHg, pulmonary wedge pressure ≤ 15mmHg and pulmonary vascular resistance ≥ 3 woods units.[1] While the pathobiology of PAH is incompletely understood, endothelial cell (EC) dysfunction has long been recognized to contribute to increased vascular tone and remodeling.[2] More recently, EC injury and apoptosis has been recognized as a central trigger for initiation of disease,[3] and underlying mutations in BMPR2, which represent the most common cause of hereditary PAH,[4] have been shown to increase susceptibility of ECs to injury and apoptosis.[5] However, it is still uncertain how EC apoptosis leads to loss of vasculature and the development of complex arterial remodeling, i.e. plexiform lesions,[3] which is a hallmark feature of advanced PAH. It has been reported that perturbed pulmonary hemodynamics are necessary for the development of occlusive arterial remodeling,[6] possibly as a result of increased intimal shear stress within the lung arteriolar bed which could contribute to ongoing endothelial injury.

We have suggested that there are distinct phases during the development of PAH;[3] the first representing the initial response to the triggering injury, which is followed by the progressive damage and loss of precapillary arterioles progressing to complex arterial remodeling at later stages of disease, largely driven by hemodynamic consequences of arteriolar pruning. Single-cell transcriptomics is a powerful tool that can elucidate the transcriptional changes underlying the onset and progression of disease at the resolution of individual cells, thereby providing new insights into the mechanisms driving the functional and structural vascular abnormalities of

PAH. Single-cell RNA sequencing (scRNA-seq) has been widely employed to better characterize lung cell populations under homeostatic conditions,[7,8] and in various disease states.[9,10] Several studies have recently employed scRNA-seq to characterize the advanced stage in several PAH experimental models.[11–14] However, all of these studies have been limited by primarily focusing on established disease, whereas the mechanisms leading to the development of PAH may be quite different in the early stages. Multiplexed scRNA-seq analysis can increase the number of samples analyzed at one time, making it feasible to study multiple timepoints with increased biological replicates by reducing costs.[15]

Therefore, we sought to employ multiplexed scRNA-seq to map out the temporal changes in transcriptomic profiles and cell populations during the full course of PAH onset and progression in the SU5416-Chronic Hypoxia (SU/CH) rat model, focusing mainly on the vascular cell populations. We were able to identify two main phases of disease progression characterized in the early stages (i.e., 1 week post SU) by widespread changes in endothelial, stromal, and immune cell populations, and the emergence of a dedifferentiated capillary (dCap) cluster, which may be undergoing endothelial to mesenchymal transition (EndoMT) based on RNA velocity analysis. At later stages of disease (i.e., 5 and 8 weeks), there was a remarkable normalization of global transcriptional activity in most lung cell populations, with the notable exception of the activated arterial ECs (aAECs), which would be most affected by perturbed lung hemodynamics, and these continued to exhibit persistent and robust changes in gene expression. Interestingly, these cells were highly localized to regions of complex arteriolar

remodeling, along with the smooth muscle (SM)-like pericytes, both expressing *Tm4sf1*, a marker previously implicated in cancer biology.

Methods

Sugen – chronic hypoxia model of PAH

All animal experiments were approved by the University of Ottawa Animal Care Committee and conducted according to the guidelines from the Canadian Council for Animal Care. Male Sprague Dawley (SD) rats (6-10 weeks old) were administered with a single dose of Sugen 5416 (SU) (20mg/kg, Tocris) delivered in CMC vehicle (as previously described)[16] or DMSO and placed in hypoxia chambers (10% O_2, Biospherix) for three weeks, followed by 5 weeks of normoxia. Animals at baseline, 1, 3, 5, and 8-weeks (**Figure 1A**) were anaesthetized by an i.p. injection of ketamine (100 mg/kg) and xyalzine (10 mg/kg) and right ventricular systolic pressure (RVSP) was measured as previously described.[17] Lungs were either isolated for tissue digestion (described below) or formalin fixed and paraffin embedded for subsequent histological analysis. The heart was excised and right ventricular hypertrophy (RVH) was assessed by the ratio of right ventricle (RV) to the left ventricle plus septum (LV+S), as previously described.[17]

Sample preparation for micro-computed tomography (microCT)

Lungs were prepared for microCT, as previously described.[18] In brief, animals were thoroughly perfused with heparinized saline (LEO Pharma Inc.) through the jugular vein with a closed chest and mechanical ventilation, using pressure matching the measured RVSP. Subsequently, chests

were opened, and an additional perfusion of heparinized saline cleared residual blood from the lungs. Barium – gelatin mix (E-Z-EM Canada Inc; Sigma-Aldrich) was perfused until consistent pressure matching RVSP was obtained. Lungs were cooled to 4°C and inflated with formalin through the trachea. Lung samples were stored at 4°C in formalin for 2 days, washed with PBS, and stored in ethanol until images were acquired with desktop microCT (SkyScan 1272, Bruker microCT). Image analysis was performed with CTAn software (Bruker microCT), and CTVox software (Bruker microCT).

H&E staining and vascular occlusion scoring

Hematoxylin and eosin (H&E) staining was performed on paraffin sections. Sections (5µm thick) were deparaffinized and rehydrated as previously described using a series of xylene, ethanol, and water washes. Sections were stained with hematoxylin for up to 8 min and washed with water. Then incubated in Scott's solution for 2 min, washed and soaked in eosin stain for 3 min. Slides were washed in water, and then dehydrated in a reverse series of ethanol and xylene rinses. Slides were mounted with Cytoseal XYL media (Thomas Scientific) and left to dry overnight. Slides were scanned with Panoramic Desk (3D Histech, Budapest, Hungary) and analyzed with CaseViewer (3D Histech).

Lung digestion and single cell preparation

Lung digestion was performed in a similar manner as previously described.[19] In brief, animals were perfused with 20ml of 0.9% saline containing 25U/ml heparin (LEO Pharm Inc.) through

the right ventricle, to drain the blood from the lungs. Lung tissue was isolated, minced with scissors and placed in Hank's Balanced Salt Solution (HBSS, Gibco) on ice. Enzymatic digestion was performed with 2500U Collagenase (Worthington), 30U Neutral Protease (Worthington), and 165U DNase I (Sigma Aldrich) in HBSS, samples were loaded into gentleMACS C tubes (Milltenyi Biotec), and placed on the OctoMACS dissociators (Milltenyi Biotec) using program 37C_LDK_m_1 for ~30min. Samples were immediately placed on ice, filtered through 70µm cell strainers (Falcon), and washed with phosphate buffered saline (PBS, Gibco) with 2mM EDTA (ThermoFischer Scientific) and 0.5% bovine serum albumin (BSA, Wisent) (PEB). Dissociated cell samples were spun at 400g for 5min at 4°C and washed again. Samples were incubated in 1x red blood cell lysis (eBioscience) for 3min, diluted with PEB, and filtered again through 70µm cell strainers (Milltenyi Biotec). Cells were washed, as above and counted with Trypan Blue (ThermoFisher Scientific) on a Cell Drop BF (DeNovix).

Multiplex single cell sample processing

Individual biological samples were barcoded using the MULTI-seq protocol described by McGinnis et al.[15] In brief, single cells were incubated with 1:1 molar ratio of anchor: unique barcode oligonucleotide sequence for 10 min at room temperature with gentle mixing. Followed by addition of the co-anchor to stabilize barcodes within membranes which was incubated for 5min on ice. Barcoded cells were washed in PBS, counted as above, and pooled at an equal ratio of cells per biological replicates. Samples with viability > 80% were sent for further processing. Pooled samples were processed using 10x Genomics Chromium (10x Genomics).

Processing of single cell RNA sequencing libraries

RNA library construction was performed using 10x Genomics Single Cell 3' RNA sequencing kit as previously described.[19] This experiment was performed with unique biological replicates on different dates, and data was later merged. Libraries were created and sequenced with NextSeq500 (Illumina) for an estimated 37554 cells with 21879 mean reads per cell, 2372 median UMI per cell, and 1118 median genes per cell (first experiment) and 35546 cells with 22322 mean reads per cell, 3152 median UMI per cell, and 1370 median genes per cell (experiment repeat). CellRanger (10x Genomics, version 6.1.2) was used to process raw sequencing reads using the Ensembl104 rat transcriptome annotations with additional inclusion of Pecam.

Quality control and single cell data analysis

Filtered feature barcode matrices were imported into R package Seurat (https://github.com/satijalab/seurat) for subsequent analysis.[20] Quality control was performed within Seurat to remove cells with >30% mitochondrial transcripts and cells with low complexity (<200 detected genes). Barcodes were demultiplexed using a manual threshold method outlined within the deMULTIplex R package (https://github.com/chris-mcginnis-ucsf/MULTI-seq).[15] Cells with multiple barcodes (doublets) or with no identifiable barcodes were removed from analysis. An additional round of doublet removal was performed with scDblFinder (https://github.com/plger/scDblFinder).[21] A total of 23,122 cells from experiment 1 and 26,611 cells from experiment 2 were retained. Data was merged into a single Seurat object for

downstream analysis. To ensure that clustering would not be impacted by batch effects or biological variability we used the integration method implemented by Seurat v3. We integrated based on biological replicates using the SCT method to regress out cell cycle and mitochondrial associated genes prior to calculation of principal component analysis (PCA) and uniform manifold approximation projection (UMAP). Cell clusters were characterized by assessment of cell-specific genes and comparison with cell classification tool Single Cell Net.[22] Cell prioritization was performed with the R package Augur (https://github.com/neurorestore/Augur) to identify populations that were most affected by PAH.[23]

Clusters corresponding to endothelial, stromal, immune, and epithelial cells were identified by *Cldn5, Col1a1, Ptprc, and Epcam*. These populations were re-clustered using the same normalization and integration approach to refine cluster identification and for subsequent analysis. Contaminating clusters were removed representing small populations with mixed canonical markers (eg. *Ptprc*+ cluster within the endothelial subset) likely representative of residual doublets not identified during previous steps.

Differential gene expression and gene set enrichment analysis

Differential gene expression analysis was performed using the R package *muscat* (https://github.com/HelenaLC/muscat) to account for the available biological replicates,[24] using standard workflow. Significant genes were identified with an adjusted p value <0.05 and a detection rate of at least 5% within the tested conditions. Gene set enrichment analysis was

performed using R package *fgsea* (version 1.20.0) on the fold change ranked list of genes for a given condition tested. Gene sets were queried and annotated which comprised all GO terms, KEGG pathways, Reactome pathways, and the MSigDB Hallmark gene sets, which were acquired from the Molecular Signatures Database (v6).[25,26] Significantly affected gene sets were filtered on an adjusted p-value of <0.05. The normalized enrichment score (NES) indicated the degree of up or down regulation of a given gene set. Volcano plots of DEGs were generated with r package EnhancedVolcano (version 1.12.0).

Ingenuity pathway analysis

Pathway enrichment analysis was performed assessing only DEGs using Ingenuity Pathway Analysis (IPA) web-based software application (Qiagen). Standard workflows were followed within the core analysis function to identify gene sets of disease or biofunctions from the IPA library that were significantly enriched based on the input DEGs.[27]

Transcription factor analysis

Transcription factor (TF) activity was predicted from available transcriptomic data using deCoupleR (version 2.0.1).[28] We employed the weighted mean method using our gene expression data to infer TF activity and their targets from the DoRothEA curated network.[29]

RNA velocity and trajectory inference analysis

We used scVelo[30] as an input for CellRank[31] to evaluate RNA velocity profiles of the combined vascular cells (ECs and stromal cells). We employed scVelo's dynamical model to infer

macrostates as an input for CellRank, which then performed RNA velocity and trajectory inferences.

Immunofluorescent staining

OCT prepared lung sections were removed from -80°C freezer and thawed to room temperature and fixed in 4% paraformaldehyde solution for 20 minutes. Sections were washed three times in 1x PBS (15 minutes per wash) and then blocked with 5% goat serum in 0.5% Triton X-100/PBS (PBS-Tx) for one hour at room temperature. After blocking, samples were incubated in primary antibodies overnight at 4°C. The next morning samples were washed in PBS-Tx (3x, 15 minutes per wash) and incubated with secondary antibodies overnight at 4°C. The following morning samples were washed in PBS-Tx and prepared for confocal microscopy with mounting media containing DAPI. All images were captured using a laser scanning Zeiss 880 confocal microscope with Fast Airyscan (Zeiss) and processed with Aivia software. Primary antibodies with their respective concentrations were as follows Griffonia Simplicifolia Lectin I (GSL I) Isolectin B4 (1:25, Vector Laboratories, #FL-1201), mouse monoclonal anti-3G5 IgM (1:10, from Ke Yuan lab), FITC anti- αSmooth Muscle Actin mouse monoclonal antibody (1:300, Sigma Aldrich, #F3777), and rabbit anti-TM4SF1 antibody (1:25, Invitrogen, #PA5-21119).

Human sample preparation

PAH patient and control lung samples were available from the Quebec Respiratory Health Network tissue bank (www.rsr-qc.ca). Paraffin sections (5µm thick) were deparaffinized and rehydrated as previously described using a series of xylene, ethanol, and water washes. Heat

mediated antigen retrieval was performed by pressure cooking in citric acid-based antigen unmasking solution. Sections were washed in PBS with 0.1% Tween-20 (PBS-T) and blocked in 5% goat serum (Cedarlane #CL1200-500) in PBS-T for 1h at room temperature. Primary antibodies were incubated overnight at 4°C using TM4SF1 (ThermoFisher Scientific #PA5-21119, 1:200) and CD31 (Dako #M0823, 1:100) antibodies. The following day samples were washed 4x in PBS-T and incubated with secondary antibodies anti-Rabbit alexa fluor 488 (ThermoFisher Scientific #A11008, 1:500) and anti-mouse alexa fluor 594 (ThermoFisher Scientific #A11005, 1:500) for 1h at room temperature. Samples were washed in PBS-T, mounted using DAPI (4',6-diamidino-2-phenylindol) Fluoromount G mounting medium (Electron Microscopy Science) and imaged with an Axio Observer microscope (Zeiss).

Statistical analysis

Data are presented as means +/- SEM. Non single cell statistical analyses were performed with GraphPad Prism v9 (GraphPad Software). Analysis of variance (ANOVA) was performed with multiple comparisons using Tukey correction used to identify statistically significant differences. P-values <0.05 were considered significant. For scRNA-seq experiments we used four animals per time point and statistical analysis was performed in line with recommendations for each analytical package used.

Results

Characterization of functional and structural vascular changes in the SU/CH PAH model

The SU/CH model exhibits many of the hemodynamic and pathological changes characteristic of human PAH, including complex arterial remodeling leading to the development of occlusive arterial lesions.[32] To better define the stages in the development of PH, we evaluated animals at serial timepoints after the administration of SU and initiation of CH (**Figure 1A**). Significant elevation of RVSP was observed as early as 1-week (47mmHg; $p = 0.056$), plateauing at >100mmHg by 5-weeks (**Figure 1B**). Right ventricular hypertrophy (RVH) was increased by 3-weeks, persisting to 8 weeks (Fulton index >60%; **Figure 1C**). In small arterioles (<50µm), there was evidence of severe arteriolar remodeling (>50% lumen loss) by five and eight weeks, compared to healthy lungs (**Figure 1D**). MicroCT analysis revealed a progressive loss of pulmonary arterial vasculature during PH progression, (**Figure 1E – H; Sup Figure 1**). The loss of the functional arterial bed primarily affected vessels <150µm in diameter, with evidence of significant arterial pruning as early as 1-week (**Figure 1F**), corresponding to the initial increase in arterial pressures, and reaching a maximal loss of 65% at 5-weeks in vessels <200µm in diameter (**Figure 1G**). RVSP and lung arterial volume were negatively correlated (**Figure 1H**). Therefore, the rat SU/CH model is well suited to explore the cellular and molecular mechanisms during the development of PAH.

Single-cell transcriptomic profiling identifies changes in endothelial and pericyte populations in PAH

Multiplexed single cell analysis was performed on lung samples from healthy controls, and 1-, 3-, 5-, 8-weeks post SU/CH initiation, incorporating four biological replicates per timepoint. Poor quality cells and those lacking a sample barcode or containing multiple barcodes were removed (**Sup Figure 2**). Additional doublet removal was performed with scDblFinder,[21] and independent experiments (23122 cells Expt1 and 26611 cells Expt2) were merged, integrated, and aligned (**Sup Figure 3A, B**). Broad cell type annotations (endothelial, stromal, myeloid, lymphoid, and epithelial) were given to resulting clusters based on the expression of canonical markers and each were then subclustered for high resolution analysis (**Sup Figure 3C**). This resulted in the classification of 44 unique lung populations (**Figure 2A**). Global UMAPs of all lung cell populations at each time point are presented in **Sup Figure 4**. At 1 week, phenotypic changes were clearly evident in many cell populations, involving all lineages, but most evident in endothelial and myeloid cells (**Figure 2B, left)**. Surprisingly, at 5 weeks there was a remarkable and somewhat paradoxical 'normalization' of global transcriptional profiles, at a timepoint when the hemodynamic and vascular changes of PAH were reaching their peak (**Figure 2B, right**). Augur[23] was used to identify cell populations most affected by SU/CH at 1- and 5-weeks relative to control cells. At 1 week, many cell clusters showed marked transcriptional changes, with aAECs, and a SM-like (*Acta2*+) pericyte population demonstrating the most substantial changes from control conditions (**Figure 2C**). In contrast, at 5 weeks, there was a reduction in these transcriptional changes with the notable exception of the aAECs (**Figure 2D**). Increases in relative cell numbers were seen in several populations including some monocytic clusters (alveolar, interstitial, and transitional macrophages, and neutrophils), both the arterial and aAEC clusters, and lymphatic ECs clusters (**Figure 2E; Sup Figures 5-7**).

However, the greatest relative increase in cell number was seen in the SM-like (*Acta2*+) pericytes beginning at 3 weeks post SU/CH (**Figure 2F).** Since the endothelial and stromal populations showed the most dramatic overall changes in the global analyses, we studied these populations in greater detail, although an overview of the myeloid, lymphoid, and epithelial cells is provided in **Sup Figures 8-10**.

Emergence of disease specific endothelial populations during PAH progression

Sub-clustering the endothelial cells produced 8 distinct cell clusters encompassing all expected populations including arterial, venous, lymphatic, general capillary (gCap) and aerocytes (aCap) ECs **(Figure 3A)**, identified by expression of accepted markers such as, *Dll4*, *Slc6a2*, *Mmrn1*, *Aplnr*, *Apln*, among others[7,33,34] (**Figure 3B and C**). Interestingly, the gCap EC population was made up of three distinct clusters. In addition to the typical gCap ECs, there was a cluster characterized by expression of early response genes (ERG), such as *Fos* and *Egr*, as has previously been described.[19] Of interest, the gCap-ERG shared expression of some genes with the venous ECs (**Figure 3C**), suggesting possible zonation changes across the capillary-venous axis.[33] The other distinct gCap cluster corresponds to previously described 'inflammatory' ECs,[35] which exhibited relative reduction in expression of canonical endothelial genes (*Cdh5*, *Cldn5*) together with upregulation of genes involved in antigen presentation (*Cd74*, *RT1Da*), consistent with a partially dedifferentiated and pro-inflammatory state.

Using proliferative markers, including *Mki67*, *Cdca8*, and *Birc5*, evidence of proliferation was seen in cells mainly located in a distinct 'proliferative' node (**Figure 3D**), representing cells from

several clusters, in particular dCap ECs and aAECs (**Figure 3A**). Endothelial UMAPs for each timepoint show the major transcriptomic changes between healthy (control) and early stages (1 and 3 weeks), or the later stages (5 and 8 weeks) of PAH progression (**Figure 3E**). Phenotypic changes were most evident in the gCap at weeks 1 and 3 but returned to the baseline distribution at weeks 5 and 8. This was accompanied at 1 week by the emergence of clusters 1 (dCap) and 4 (aAECs) which largely persisted to week 8 (**Figure 3F**). Comparing fold-change with respect to the total EC population in each cluster relative to control, there was an early (1 week) and sustained increase in dCap and aAECs, with increases at later time points in arterial, and lymphatic ECs (**Figure 3G**). Again, this supports an important role for the aAECs and dCap ECs in the onset and progression of PAH.

Differential gene expression analysis in activated arterial and dedifferentiated EC populations

The aAECs and dCap ECs emerged during PAH progression in the SU/CH rat model. To better understand the distinct transcriptomic profiles within these populations, we compared changes in expression of genes between classical arterial and aAECs and between gCap and dCap ECs (including all timepoints) (**Figure 4A and B**, respectively). aAECs were characterized by reduced expression of classical arterial genes[7] including the notch ligands *Dll4* and *Notch1*, gap junction proteins *Gja5*, signaling molecules *Efnb2* and *Efna1*, and receptors *Cxcr4* (**Figure 4A**). Conversely aAECs showed increased expression of extracellular matrix (ECM) proteins fibronectin (*Fn1*), Matrix gla (*Mgp*), and elastin (*Eln*), and signaling molecule stromal derived factor 1 (*Cxcl12*), all consistent with an activated endothelial phenotype. Of particular interest was transmembrane 4L6 family 1 (*Tm4sf1*) with the greatest fold change, since this has been strongly implicated in

cancer cell growth[36] (**Figure 4A**). The differences in gene expression were slightly more modest between gCap and dCap reflecting a de-differentiated EC state compared to gCap with relative decreases in EC tight junction genes VE-cadherin (*Cdh5*), and claudin 5 (*Cldn5*), as well as classic gCap genes apelin receptor (*Aplnr*), and endomucin (*Emcn*) (**Figure 4B**). Interestingly, numerous genes associated with ribosomal proteins were upregulated by dCap, consistent with a relatively de-differentiated state,[37] together with genes associated with antigen presentation and inflammation (*Cd74*, *RT1-Da*), as well as Tumor Protein, Translationally-Controlled 1 (*Tpt1*) (**Figure 4B**), an oncogene and proinflammatory factor previously implicated in PAH.[38,39] The differences in relative expression between all EC populations are depicted in **Figure 4C**, confirming the relative overexpression of genes associated with an activated EC state in aAECs, in particular *Tm4sf1* and the predominance of antigen presenting genes and *Tpt1* in dCap, while both displayed relatively lower expression of classical EC genes, but to a greater extent in dCap. Ingenuity pathway analysis (IPA) was performed, again comparing activated arterial versus arterial ECs, and dCap versus gCap. There was a remarkable similarity in the overall pattern of gene sets that were enriched or decreased between these two distinct populations. Both aAECs and dCap ECs were enriched for disease states or biofunctions associated with metabolic processes, translation, mobility, and inflammation consistent with heightened cell growth relative to classic arterial and gCap ECs, respectively (**Figure 4D and E**). Conversely, aAECs and dCap ECs showed reduced expression of gene sets associated with angiogenesis and vascular development compared their respective comparator EC populations (**Figure 4D and E**). This is consistent with these EC populations playing a central role in vascular inflammation and arterial remodeling in the development of PAH.

Changes in stromal cells in response to PAH

Next, we re-clustered all stromal cell populations (fibroblasts, pericytes, smooth muscle cells) into 6 unique populations, including matrix fibroblasts, *Col13a1*+ fibroblasts, *Col14a1*+ fibroblasts, SM-like (*Acta2+*) pericytes, classical (*Pdzd2*+) pericytes, and smooth muscle cells (SMCs)/myofibroblasts (**Figure 5A**). Fibroblasts were identified based on *Pdgfra* and *Tcf21* expression, and differential expression of *Col13a1* and *Col14a1*; while pericytes were identified by expression of *Pdgrb*, *Csgp4,* and *Mcam* (**Figure 5B and C**).[9,40] Pericytes were further distinguished into 'classical' pericytes uniquely expressing *Pdzd2*, *Postn*, *Gucy1b1* and 'SM-like' pericytes based on expression of contractile markers (*Acta2, Des*) (**Figure 5B and C**), as previously described by Hurskainen et al.[9]. SMCs and myofibroblasts were observed in a mixed population expressing *Acta2* and *Dcn*. Interestingly, both pericyte populations showed marked expansion during PAH progression, whereas there was little change in the global proportion of other stromal cell populations (**Figure 5D and E**), suggesting a possible contribution to arteriolar remodeling in PAH.

We next compared transcriptomic profiles of classical versus SM-like pericytes (at all timepoints). The SM-like, *Acta2*+ pericytes were characterized by reduced expression of some classic pericyte genes, including *Postn* and *Vtn*. Additionally, *Gucy1a1* and *Pdzd2* appeared to predominate in 'classical' pericytes. Conversely, the SM-like pericytes exhibited high expression of a number of contractile genes including: *Acta2*, *Des*, *Myh11*, and *Mustn1* (**Sup Figure 11A and B**). Gene set enrichment analysis (GSEA) was performed comparing SM-like pericytes

versus classical pericytes. The SM-like pericytes exhibited robust enrichment of gene sets associated with contraction, muscle differentiation, and translation, while gene sets associated with cell locomotion, lipid processing, and organ development were decreased (**Sup Figure 11C**). Interestingly, SM-like (Acta2+) pericytes also expressed *Tm4sf1* (**Sup Figure 11B**), which was markedly increased in the aAECs as well and has been implicated in epithelial to mesenchymal transition (EMT).[41] Therefore, these data are consistent with the SM-like pericytes contributing to arteriolar remodeling and distal muscularization in PAH.

Endothelial transcriptomic changes during PAH progression

By performing scRNA-seq at multiple timepoints in the SU/CH model, we were able to discern changes in gene expression profiles during the development of the PAH phenotype. The greatest number of differentially expressed genes (DEGs) was seen 1-week compared to healthy controls, across numerous cell types (**Sup Figure 12**); however, this was most marked for ECs (**Figure 6A**). Interestingly, the number of DEGs in many lung cell populations normalized at later stages (5-, 8- weeks), with the notable exception of the aAECs which exhibited persistent high numbers of DEGs (**Figure 6A-C**). Three distinct patterns of gene expression could be discerned (**Figure 6D**). The first was represented by early increases in gene expression peaking at 1 week, with a partial normalization by 5 to 8 weeks. These were largely cell-cycle associated genes, including *Birc5*, *Cdc20*, *Cdkn2c*, *Cdkn3*, and *Cdk1* consistent with an increase in EC proliferation observed at 1 week (**Figure 6D**). Interestingly, a similar pattern of genes was observed in the dCap ECs(**Figure 6D**), consistent with the increase in EC proliferation observed at 1 week (**Figure 3D, E**). The second pattern was unique to the aAECs and was represented by

genes that showed a sustained increase in expression at all timepoints, including a number of genes which have previously been implicated in PH such as, *Ramp1*,[42] *Trpv4*,[43] *Jag1*,[44] and *Pdgfa*[45]. Finally, the third pattern was made up of genes strongly expressed by both classical arterial ECs and aAECs at baseline, and largely involved in angiogenesis and vascular development, such as *Dll4*, *Notch1*, *Efna1*, and *Erg1*, which were selectively downregulated only in aAECs during PH development (**Figure 6D**). Not unexpectedly, *Cyp1a1* and *Cyp1b1*, which mediate drug metabolism,[46] where upregulated in all EC populations after treatment with SU, with the exception of lymphatic ECs (**Figure 6D**). For a more comprehensive evaluation of the potential biological impact of these transcriptional changes, we performed ingenuity pathway analysis (IPA) of changes in gene expression in aAECs between control and PAH timepoints. This confirmed the upregulation of gene sets associated with increased cell proliferation and cancer-like cell growth, with downregulation of gene sets associated with angiogenesis, vasculogenesis and cell death in aAECS (**Figure 6E**). Together, these results point to aAECs as a unique lung EC population exhibiting a transcriptional signature consistent with a growth dysregulated state, associated with decreased reparative and angiogenic capacity, and likely playing an important role in arterial remodeling in PAH.

Spatial localization of *Tm4sf1*-expressing activated arterial ECs

Both the aAECs and SM-like (*Acta2*+) pericytes exhibited high expression of *Tm4sf1*, raising the possibility that this may be a marker of these populations. Therefore, we examined *Tm4sf1* expression across all 44 lung cell populations, and found that *Tm4sf1* was largely restricted to arterial and venous ECs, as well as pericytes and mesothelial cells, with by far the greatest

expression in aAECs (**Figure 7A**). We then used immunofluorescent staining for TM4SF1, and co-staining for CD31 (ECs), 3G5 (pericytes) or smooth muscle actin (SMA; muscularized cells), to identify the anatomic location of these cell populations in the healthy and diseased lungs. In control lung samples, TM4SF1 expression was localized mainly to $3G5^{+}$ pericytes surrounding alveolar capillary ECs, with sparse EC staining (**Figure 7B**). However, at 5 weeks post SU/CH, there was abundant TM4SF1 staining associated with ECs, as evidence by isolectin B4, localized mainly to severely remodeled arterioles, consistent with a direct role for the aAECs in occlusive arteriopathy with accumulation of TM4SF1+ ECs within the vascular lumen (**Figure 7C**). Interestingly, in the diseased lung, 3G5 staining was associated mainly with regions of arteriolar muscularization, which also stained for positive for SMA, suggesting a role for the SM-like pericytes in distal arteriolar remodeling; however, these regions of muscularization showed markedly reduced staining for TM4SF1. We also assessed the spatial distribution of TM4SF1 expression in human lung samples from control and PAH patients (**Figure 7D**). In the control samples, TM4SF1 staining could be seen surrounding small arterioles, but was not associated with ECs (CD31 staining). In contrast, in PAH samples there was abundant TM4SF1 staining, mainly localized to the perivascular region but also adjacent to the intima, and clearly associated with ECs, with a relative paucity of staining in the intervening SMC layer (**Figure 7D**). Together these results confirm that TM4SF1 positive ECs and perivascular cells are associated with complex arterial remodeling in PAH, consistent with a pathological role of aAECs and SM-like pericytes in this disease.

RNA velocity of endothelial and stromal cells during SU/CH progression

Endothelial to mesenchymal transition (EndoMT) has been previously implicated in PAH.[47] To explore a potential role of cell fate transitions in the development of PAH we employed RNA velocity analysis, using scVelo dynamical modelling,[30] specifically addressing population dynamics within the EC and stromal clusters based on their importance in arterial remodeling. Surprisingly, RNA velocity analysis demonstrated strong vectors within the dCap ECs cluster, leading to fibroblast populations (**Figure 8A**), consistent with the possibility of EndoMT contributing to vascular stromal cells in this PAH model. Strong velocity vectors were also observed from the classical (*Pdzd2*+) pericytes towards the SM-like (*Acta2*+) pericytes (**Figure 8A**), consistent with pericyte differentiation towards a smooth-muscle cell fate as a mechanism of distal muscularization. Interestingly, the were also vectors from the gCap-ERG cluster leading to the venous ECs, raising the possibility of zonation of this cluster towards the venous side of the alveolar capillary. The possibility of EndoMT was further supported by a latent time analysis which provides an estimate of a cell's internal clock, from progenitor fates (approaching 0) to terminally differentiated cells (approaching 1). This model identified aCap and gCap ECs, *Col13a1*$^{+}$ fibroblasts, and classical (*Pdzd2*$^{+}$) pericytes as being close to 'terminal' states, while aAECs, dCap ECs, and SM-like pericytes were closer to a progenitor state (**Figure 8B**). We then inferred transcription factor (TF) activities from gene expression data for the EC clusters using deCoupleR.[28] TFs classically associated with EMT[48] including *Zeb1*, *Zeb2*, *Snai2* had predicted activity in dCap EC cluster, as well as some other EC populations, whereas *Twist1* was seen mainly in gCap-ERG ECs (**Figure 8C**). The top 40 most variable TFs are presented in **Figure 8D**. Interestingly, gCap-ERG and venous ECs shared a number of TFs (**Figure 8D**), further supporting zonation between these populations based on similarities in gene expression as shown in **Figure**

3C. As well, dCap ECs exhibited elevated activity associated with *Nr2f2*[49] and *Zeb2,*[48] genes which have been implicated in EndoMT, further supporting a potential role for this dedifferentiated EC population in EndoMT during PAH progression.

Discussion

PAH is characterized by a marked pruning of the pulmonary arteriolar bed which, together with complex arterial remodeling, leads to progressive increases in pulmonary vascular resistance; yet the mechanisms underlying these dramatic vascular changes remains obscure. The rat SU/CH model of PAH provides an ideal opportunity to study functional, structural, and molecular changes during the onset of PH and the development of complex arterial remodeling, something which cannot be accomplished in PAH patients who usually present late and typically lung tissue is only available at the very end stage of disease. Using multiplexed scRNA-seq we analyzed four timepoints in the SU/CH model, spanning early to late disease, compared to healthy control animals. Major transcriptomic changes were observed in nearly all lung cell populations at 1-week post SU/CH, but paradoxically at later timepoints, there was substantial normalization in gene expression profiles of nearly all lung cell populations including immune, stromal, and vascular cells, with the notable exception of an 'activated' arterial EC cluster (aAECs). This finding may be surprising considering that the pathological and hemodynamic changes are minimal at 1 week in this model, and severe PH only manifests at later time points. However, the marked and widespread early transcriptional changes likely reflect an injury-repair response induced by administration of the VEGFR2 antagonist coupled with hypoxia,

whereas the persistent changes in aAECs suggest that this population, in particular, may play a central role in the onset and progression of vascular disease.

Since the injurious effects of SU and hypoxia are no longer operative at these late timepoints, this begs the question of what induces the persistent and marked transcriptomic response selectively in the aAEC cluster. It has been reported that complex arterial remodeling in this model is dependent on perturbed hemodynamics,[6] likely due to pathological levels of shear stress, which would have the greatest impact on the distal arteriolar bed.[50] Moreover complex lesion formation could be prevented and reversed by unilateral flow-limiting pulmonary arterial banding.[6] Using microCT, we demonstrated a progressive loss of arteriolar volume in the SU/CH model, which was significant as early as 1 week, at timepoint that preceded any evidence of occlusive arterial remodeling. This suggests that the early arteriolar loss, probably as a direct consequence of widespread EC apoptosis, was a major cause for the initial hemodynamic changes in this model. It should also be recognized that in a dispersed lung cell preparation, arterial ECs are predominantly derived from distal lung arterioles, since the surface area, and hence number of ECs, is the greatest in the microcirculation. Thus, aAECs represent the lung arterial endothelial population that would be the most likely to be impacted by hemodynamic perturbations in developing PH, which could provide the stimulus for the persistent transcriptomic changes observed in this EC subset.

At the early timepoint (1 week), aAECs and dCap ECs exhibited a transcriptional signature characterized by the transient expression of genes related to cell proliferation and repair.

However, at later time points (5 and 8 weeks), only aAECs exhibited a persistent gene expression profile that was consistent with dysregulated cell growth and reduced capacity for vascular repair and angiogenesis. This is in close agreement with a broad literature showing the emergence of hyperproliferative and apoptosis-resistant ECs in PAH,[51] that are believed to result in occlusive arterial remodeling. We have also demonstrated that aAECs can be distinguished from other EC populations, including classical arterial ECs, by high expression of a trans-membrane protein, TM4SF1, which is also found in many cancer cells.[36] Interestingly, this marker was also expressed by the SM-like, *Acta*2+ pericytes. Using immunofluorescence staining, we were able to show that TM4SF1^{+} ECs were abundant in complex arterial lesions in the SU/CH model and as well as in lung sections from patients with PAH, particularly in regions of occlusive arteriopathy, consistent with a dominant role for this EC population arterial remodeling. Interestingly, while TM4SF1 staining was also associated with pericytes surrounding alveolar capillaries in the healthy lung, in the diseased lung 3G5^{+} pericytes within muscularized regions exhibited reduced staining for this marker, consistent with further differentiation to a more mature SMC phenotype. This finding suggests that the SM-like pericyte cluster, which by RNA velocity analysis originated from the classical Pdzd2^{+} pericytes, is likely responsible for the marked distal muscularization of arterioles in PAH, as has been previously suggested.[52,53]

Hong et al. previously identified a similar subset of ECs expressing *Tm4sf1*[54] in both the rat monocrotaline (MCT) model and SU/CH model in late-stage PH.[11] However, based on their transcriptomic profile, this *Tm4sf1*$^{+}$ EC population may have been heterogenous, containing

capillary, arterial, and venous ECs. While we saw some *Tm4sf1* expression in venous and classical arterial ECs, this was much less than in the aAECs, which also demonstrated a transcriptomic signature of dysregulated cell growth. While TM4SF1 may be a useful marker for aAECs, it is also possible that it may contribute to growth dysregulation. TM4SF1 is a trans-membrane protein belonging to the tetraspanin superfamily that regulates numerous signaling pathways modulating cell development, activation, growth, and motility.[36] Increased *Tm4sf1* expression has been observed in ECs in cancer, as well as in malignant cells, particularly in breast, lung and ovarian cancers,[36] and is a promising target for novel cancer therapies.[55–57] Therefore, similar strategies to antagonize or silence *Tm4sf1* may have benefits in reducing EC growth and arterial remodeling in PAH.

We also identified a second EC population in the PAH model, termed dCap ECs, which exhibited a relatively de-dedifferentiated cell state with reduced expression reduced of tight junction related genes such as *Cdhn5* and *Cldn5*, indicative of a loss of functional integrity, along with upregulation of antigen presentation related genes such as *Cd74* and *RT1-Da*. A similar 'proinflammatory' EC population has been previously implicated in PAH identified by higher expression of CD74.[35] As a receptor for macrophage migratory inhibitory factor (MIF), CD74 may contribute to inflammatory cell recruitment which is well known to be associated with arterial remodeling in PAH.[36] Enriched expression of *Cd74* was also described in a single cell transcriptomic analysis of lung ECs isolated from a SU/CH mouse of PH;[12] however, *Cd74* expression was increased in multiple EC populations including capillary, arterial, and venous. As well, this model does not exhibit occlusive arterial remodeling and the authors found no

evidence of EC proliferation in their single cell transcriptomic analysis.[12] In our study, we were able to determine that *Cd74* enriched ECs represented a discrete subset of gCap ECs that were characterized by a dedifferentiated state. As well we demonstrated a clear transcriptomic signature for cell proliferation, which was portrayed in the UMAP as a 'proliferative node' consisting of cells from several clusters, including the aAECs and the *Cd74* enriched dCap ECs. This population also exhibited a profound early expansion at 1-week in the SU/CH rat model, persisting throughout PAH development.

While the precise role of this dedifferentiated EC population in the pathogenesis of PAH is unclear, our study provides novel evidence that it may contribute to EndoMT, thereby giving rise to stromal cells that have been long implicated in arterial remodeling.[47] EndoMT shares many similarities with epithelial to mesenchymal transition (EMT), and is a phenomenon characterized loss of canonical gene expression associated with functional integrity, such as cell-cell junctions, and the adoption of motile and invasive behaviors.[48] This intriguing possibility was supported by RNA velocity analysis, which showed numerous strong vectors originating in the dCap EC cluster and directed to fibroblast populations. Moreover, latent time analysis showed that dCap ECs approximated a progenitor cell fate whereas other gCap EC populations were closer to a 'terminal' state of cell differentiation. While were able to infer increased activity of some classical TFs implicated in EndoMT in this population, such as *Zeb1* and *Snai2*, these changes were subtle. However, increasingly it has been recognized that EMT is a complex process and that EndoMT status cannot be determined based on a small number of molecular markers.[48] Rather, EndoMT is supported by loss of typical endothelial characteristics

associated with functional integrity (i.e. tight junction proteins like cadherins and claudins), and adoption of an invasive phenotype with increased motility. Indeed, dCap ECs exhibited a more transcriptionally active, progenitor-like state as identified through DEGs and IPA analysis and showed increased expression of several other genes related to EndoMT and cancer-like cell mobility, including *Tpt1*,[38] *Crip1*,[58] and *S100a4*.[59]

Therefore, we have used serial single cell transcriptomics to identify two discreet EC populations that appear to play a dominant role in arterial remodeling during the progression of PAH. While it is likely that these correspond to populations previously identified in PAH, our study provides important new insights into their specific roles in the pathobiology of PAH. Notably we show that *Cd74*-enriched ECs, previously described as pro-inflammatory, represent a distinct gCap EC population also exhibiting features of dedifferentiation and loss of functional integrity (i.e., dCap ECs), that may participate in EndoMT, thereby contributing to the generation of profibrotic arterial fibroblasts and other stromal cells that also play an important in complex lesion formation in PAH. As well, we have shown that an activated arterial EC cluster (i.e., aAECs) adopts a marked growth dysregulated transcriptional state, likely in response to ongoing hemodynamic perturbations that predominantly impact the distal arteriolar bed. For the first time we provide evidence that these cells are spatially localized to remodeled arterioles, particularly in regions of intimal obliteration, in both the SU/CH model and PAH patients. Also, we show that the activated arterial ECs are closely associated with a SM-like pericytes derived from classical pericytes, which contribute directly to distal arteriolar muscularization in remodeled arterioles. Finally, since both cell types can be identified by

Tm4sf1 expression, this may be a promising therapeutic target for preventing or reversing arterial remodeling in PAH.

Acknowledgements
We would like to thank Anli Yang for her technical assistance with all animal experiments and Elmira Safaie Qamsari for her assistant sectioning fresh frozen tissues for immunofluorescent staining. Additionally, we thank the following Ottawa Hospital Research Institute core services Stemcore laboratories for single cell library preparation and sequencing, OHRI Bioinformatics core for preprocessing of single cell data, and the University of Ottawa's Animal Care and Veterinary Services core.

Sources of Funding
This work was supported by a Foundation award from the Canadian Institutes of Health Research (FDN – 143291) to DJS. NDC acknowledges scholarship support from the Canadian Institutes of Health Research, and the Canadian Vascular Network.

Competing Interests
None

Author Contributions
NDC has contributed to all aspects of this work including conception, experimental design, performing the experiments, data analysis and preparation of this manuscript. EM contributed to data analysis and figure preparation. RSG assisted with collection of the first experiment of single cell samples and preliminary analysis. YD was instrumental in preparation and collection of the lung samples for microCT analysis and in all animal studies. KS performed the IPA analysis. AS performed the RNA velocity analysis and contributed to figure preparation. DPC assisted with single cell transcriptomic analysis and provided valuable support in initial experimental design. SEL performed immunofluorescent staining on all human samples and contributed to figure preparation. TK performed immunofluorescent staining on fresh frozen rat sections. KY collected images on rat lung sections, assisted with figure preparation and data interpretation, and manuscript review. SB provided access to human PAH samples, contributed expertise to interpretation of results, and manuscript review. DJS was involved in the conception, experimental design, data analysis and interpretation, figure design, and drafting the manuscript. DJS is accountable for all aspects of this work.

References

1. Simonneau, G. *et al.* Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J* **53**, (2019).
2. Evans, C. E., Cober, N. D., Dai, Z., Stewart, D. J. & Zhao, Y.-Y. Endothelial Cells in the Pathogenesis of Pulmonary Arterial Hypertension. *European Respiratory Journal* 2003957 (2021) doi:10.1183/13993003.03957-2020.
3. Cober, N. D., VandenBroek, M. M., Ormiston, M. L. & Stewart, D. J. Evolving Concepts in Endothelial Pathobiology of Pulmonary Arterial Hypertension. *Hypertension* **79**, 1580–1590 (2022).
4. Deng, Z. *et al.* Familial primary pulmonary hypertension (gene PPH1) is caused by mutations in the bone morphogenetic protein receptor-II gene. *Am J Hum Genet* **67**, 737–744 (2000).
5. Teichert-Kuliszewska, K. *et al.* Bone morphogenetic protein receptor-2 signaling promotes pulmonary arterial endothelial cell survival: Implications for loss-of-function mutations in the pathogenesis of pulmonary hypertension. *Circ Res* **98**, 209–217 (2006).
6. Abe, K. *et al.* Haemodynamic unloading reverses occlusive vascular lesions in severe pulmonary hypertension. *Cardiovasc Res* **111**, 16–25 (2016).
7. Schupp, J. C. *et al.* Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung. *Circulation* **144**, 286–302 (2021).
8. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
9. Hurskainen, M. *et al.* Single cell transcriptomic analysis of murine lung development on hyperoxia-induced damage. *Nature Communications 2021 12:1* **12**, 1–19 (2021).
10. Sikkema, L. *et al.* An integrated cell atlas of the lung in health and disease. *Nature Medicine 2023 29:6* **29**, 1563–1577 (2023).
11. Hong, J. *et al.* 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* **203**, 1006–1022 (2021).
12. Rodor, J. *et al.* Single-cell RNA-seq profiling of mouse endothelial cells in response to pulmonary arterial hypertension. *Cardiovasc Res* (2021) doi:10.1093/cvr/cvab296.
13. Asosingh, K. *et al.* Single-cell transcriptomic profile of human pulmonary artery endothelial cells in health and pulmonary arterial hypertension. *Scientific Reports 2021 11:1* **11**, 1–13 (2021).
14. Crnkovic, S. *et al.* Single-cell transcriptomics reveals skewed cellular communication and phenotypic shift in pulmonary artery remodeling. *JCI Insight* **7**, (2022).
15. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* **16**, 619–626 (2019).
16. Taraseviciene-Stewart, L. *et al.* Inhibition of the VEGF receptor 2 combined with chronic hypoxia causes cell death-dependent pulmonary endothelial cell proliferation and severe pulmonary hypertension. *The FASEB Journal* **15**, 427–438 (2001).
17. Chaudhary, K. R. *et al.* Efficacy of treprostinil in the SU5416-hypoxia model of severe pulmonary arterial hypertension: haemodynamic benefits are not associated with improvements in arterial remodelling. *Br J Pharmacol* **175**, 3976 (2018).

18. Deng, Y. *et al.* Optimizing imaging of the rat pulmonary microvasculature by micro-computed tomography. *Pulm Circ* **9**, (2019).
19. Godoy, R. S. *et al.* Single-cell transcriptomic atlas of lung microvascular regeneration after targeted endothelial cell ablation. *Elife* **12**, (2023).
20. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
21. Germain, P.-L. *et al.* Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research 2021 10:979* **10**, 979 (2021).
22. Tan, Y. & Cahan, P. SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species. *Cell Syst* **9**, 207-213.e2 (2019).
23. Skinnider, M. A. *et al.* Cell type prioritization in single-cell data. *Nat Biotechnol* **39**, 30–34 (2021).
24. Crowell, H. L. *et al.* muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat Commun* **11**, (2020).
25. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**, 417–425 (2015).
26. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005).
27. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).
28. Badia-i-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances* **2**, (2022).
29. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* **29**, 1363–1375 (2019).
30. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology 2020 38:12* **38**, 1408–1414 (2020).
31. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Nature Methods 2022 19:2* **19**, 159–170 (2022).
32. Stenmark, K. R., Meyrick, B., Galie, N., Mooi, W. J. & McMurtry, I. F. Animal models of pulmonary arterial hypertension: The hope for etiological discovery and pharmacological cure. *Am J Physiol Lung Cell Mol Physiol* **297**, 1013–1032 (2009).
33. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764-779.e20 (2020).
34. Wakabayashi, T. & Naito, H. Cellular heterogeneity and stem cells of vascular endothelial cells in blood vessel formation and homeostasis: Insights from single-cell RNA sequencing. *Front Cell Dev Biol* **11**, (2023).
35. Le Hiress, M. *et al.* Proinflammatory Signature of the Dysfunctional Endothelium in Pulmonary Hypertension. Role of the Macrophage Migration Inhibitory Factor/CD74 Complex. *Am J Respir Crit Care Med* **192**, 983–997 (2015).
36. Fu, F. *et al.* Role of Transmembrane 4 L Six Family 1 in the Development and Progression of Cancer. *Front Mol Biosci* **7**, (2020).

37. Kim, Y.-K. *et al.* Absolute scaling of single-cell transcriptomes identifies pervasive hypertranscription in adult stem and progenitor cells. *Cell Rep* **42**, 111978 (2023).
38. Lavoie, J. R. *et al.* Proteomic Analysis Implicates Translationally Controlled Tumor Protein as a Novel Mediator of Occlusive Vascular Remodeling in Pulmonary Arterial Hypertension. *Circulation* **129**, 2125–2135 (2014).
39. Ferrer, E. *et al.* A Potential Role for Exosomal Translationally Controlled Tumor Protein Export in Vascular Remodeling in Pulmonary Arterial Hypertension. *Am J Respir Cell Mol Biol* **59**, 467–478 (2018).
40. Muhl, L. *et al.* Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat Commun* **11**, 3953 (2020).
41. Tang, Q. *et al.* TM4SF1 promotes EMT and cancer stemness via the Wnt/β-catenin/SOX2 pathway in colorectal cancer. *Journal of Experimental & Clinical Cancer Research* **39**, 232 (2020).
42. Qing, X. & Keith, I. M. Targeted blocking of gene expression for CGRP receptors elevates pulmonary artery pressure in hypoxic rats. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **285**, L86–L96 (2003).
43. Cussac, L.-A. *et al.* TRPV4 channel mediates adventitial fibroblast activation and adventitial remodeling in pulmonary hypertension. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **318**, L135–L146 (2020).
44. Xiao, Y., Gong, D. & Wang, W. Soluble Jagged1 Inhibits Pulmonary Hypertension by Attenuating Notch Signaling. *Arterioscler Thromb Vasc Biol* **33**, 2733–2739 (2013).
45. Solinc, J. *et al.* The Platelet-Derived Growth Factor Pathway in Pulmonary Arterial Hypertension: Still an Interesting Target? *Life* **12**, 658 (2022).
46. Kwapiszewska, G., Johansen, A. K. Z., Gomez-Arroyo, J. & Voelkel, N. F. Role of the Aryl Hydrocarbon Receptor/ARNT/Cytochrome P450 System in Pulmonary Vascular Diseases. *Circ Res* **125**, 356–366 (2019).
47. Stenmark, K. R., Frid, M. & Perros, F. Endothelial-to-Mesenchymal Transition. *Circulation* **133**, 1734–1737 (2016).
48. Yang, J. *et al.* Guidelines and definitions for research on epithelial–mesenchymal transition. *Nat Rev Mol Cell Biol* **21**, 341–352 (2020).
49. Xia, B. *et al.* NR2F2 plays a major role in insulin-induced epithelial-mesenchymal transition in breast cancer cells. *BMC Cancer* **20**, 626 (2020).
50. Postles, A., Clark, A. R. & Tawhai, M. H. Dynamic blood flow and wall shear stress in pulmonary hypertensive disease. in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014* vol. 2014 5671–5674 (Institute of Electrical and Electronics Engineers Inc., 2014).
51. Masri, F. A. *et al.* Hyperproliferative apoptosis-resistant endothelial cells in idiopathic pulmonary arterial hypertension. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **293**, L548–L554 (2007).
52. Yuan, K. *et al.* Mural Cell SDF1 Signaling Is Associated with the Pathogenesis of Pulmonary Arterial Hypertension. *Am J Respir Cell Mol Biol* **62**, 747–759 (2020).
53. Bordenave, J. *et al.* Lineage Tracing Reveals the Dynamic Contribution of Pericytes to the Blood Vessel Remodeling in Pulmonary Hypertension. *Arterioscler Thromb Vasc Biol* **40**, 766–782 (2020).

54. Hong, J. *et al.* Tm4sf1-marked Endothelial Subpopulation Is Dysregulated in Pulmonary Arterial Hypertension. *Am J Respir Cell Mol Biol* (2022) doi:10.1165/RCMB.2022-0020OC.
55. Shen, Y. *et al.* Construction of CAR-T cells targeting TM4SF1 and its anti-tumor capacity in ovarian cancer. *Immunol Lett* **255**, 1–9 (2023).
56. Chen, G. *et al.* Targeting TM4SF1 exhibits therapeutic potential via inhibition of cancer stem cells. *Signal Transduct Target Ther* **7**, 350 (2022).
57. Lin, C.-I. *et al.* TM4SF1: a new vascular therapeutic target in cancer. *Angiogenesis* **17**, 897–907 (2014).
58. Zhang, L.-Z., Huang, L.-Y., Huang, A.-L., Liu, J.-X. & Yang, F. CRIP1 promotes cell migration, invasion and epithelial-mesenchymal transition of cervical cancer by activating the Wnt/β-catenin signaling pathway. *Life Sci* **207**, 420–427 (2018).
59. Gorelova, A., Berman, M. & Al Ghouleh, I. Endothelial-to-Mesenchymal Transition in Pulmonary Arterial Hypertension. *Antioxid Redox Signal* **34**, 891–914 (2021).

Figures

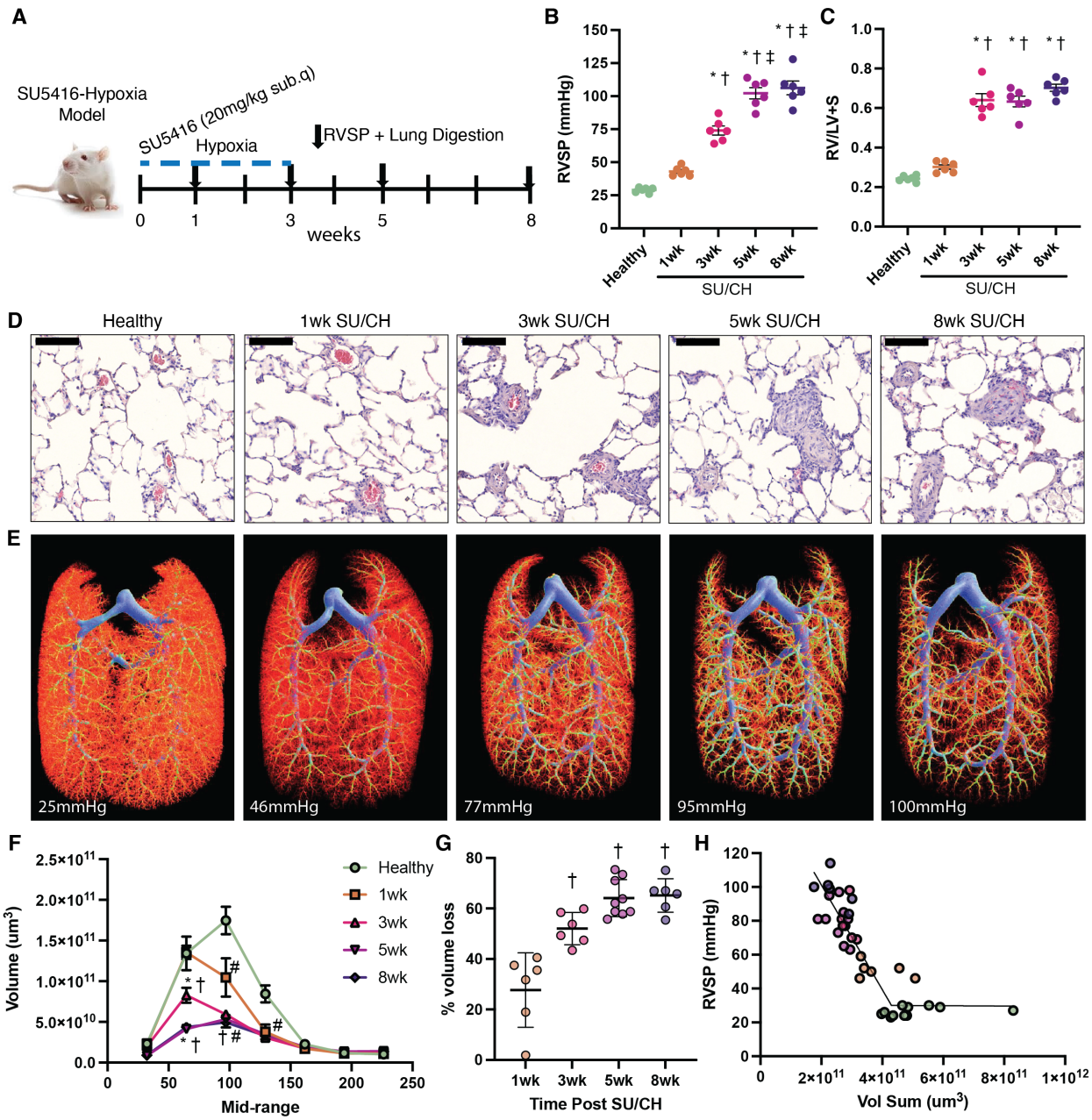

Figure 1: Characterization of Sugen-Chronic Hypoxia (SU/CH) model of PAH. A) Timeline of the SU/CH model and relevant end of study sampling. B) Right ventricular systolic pressure (RVSP). C) Right ventricular hypertrophy (RV/LV+S). D) Representative H&E histological images showing typical features of arterial remodeling associated with PH. E) Representative imaged of micro computed tomography (MicroCT) arterial angiograms of the pulmonary circulation during the development of PH. F) Summary of quantitative analysis of arterial volumes showing the greatest reduction in arteries under 200µm in diameter. G) Volume loss relative to healthy control in vessels under 200µm in diameter. H) Arterial volume (<500µm diameter) was negatively correlated with RVSP during PH development. Data represented as mean ± SEM, n =

6 – 12. * $p < 0.05$ for 3-, 5-, 8- weeks SU/CH vs control, † $p < 0.05$ for 3-, 5-, 8- weeks vs 1-week, ‡ $p < 0.05$ for 5-, 8- weeks vs 3-weeks, # $p < 0.05$ for 0.01 for all timepoints vs control. Scale bar represents 100μm.

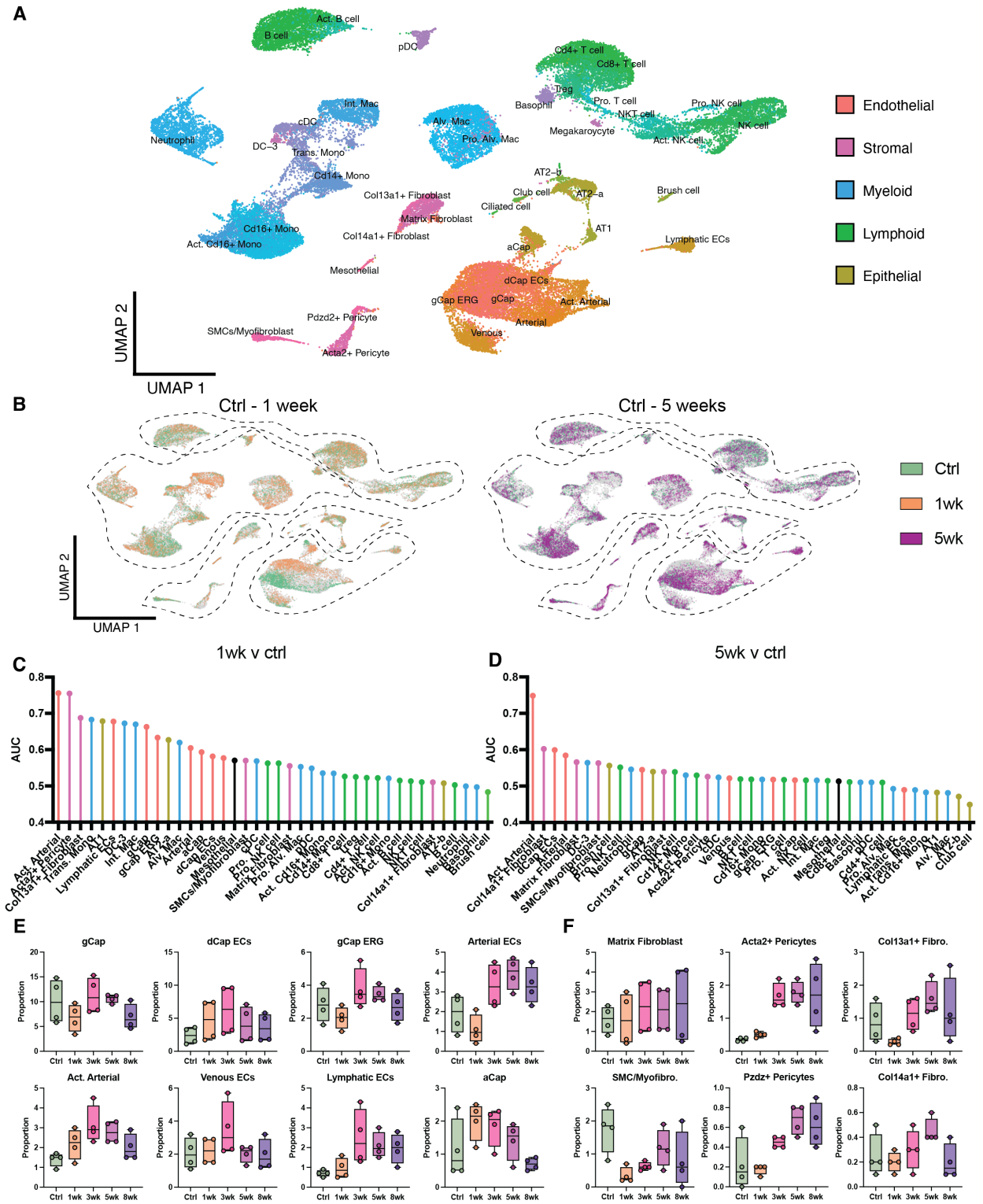

Figure 2: Multiplexed single cell transcriptomic atlas of lung cells during PAH progression. A) 44 unique lung clusters were identified and integrated into the global uniform manifold approximation and projection (UMAP). B) Global UMAPs colored by timepoint with control (green) compared with 1-week (orange) and 5-weeks (purple). Cell prioritization was performed

using a machine learning algorithm to identify clusters that were most affected between 1-week and control (C) and 5-weeks and control (D). At 1-week major transcriptomic changes in many populations were identified, while at 5-weeks, activated arterial ECs (aAECs) were the most affected cell population. Changes in global proportions of cell populations were quantified in the endothelial (E) and stromal (F) subsets by timepoint, demonstrating relative increases in arterial ECs, activated arterial ECs, *Acta2*+ pericytes, and classical pericytes during PAH progression compared to controls. Multiplexed single cell data was obtained from n = 4 animals per timepoint, multiplexed using unique barcodes, pooled and subjected to library construction using 10x-Genomics.

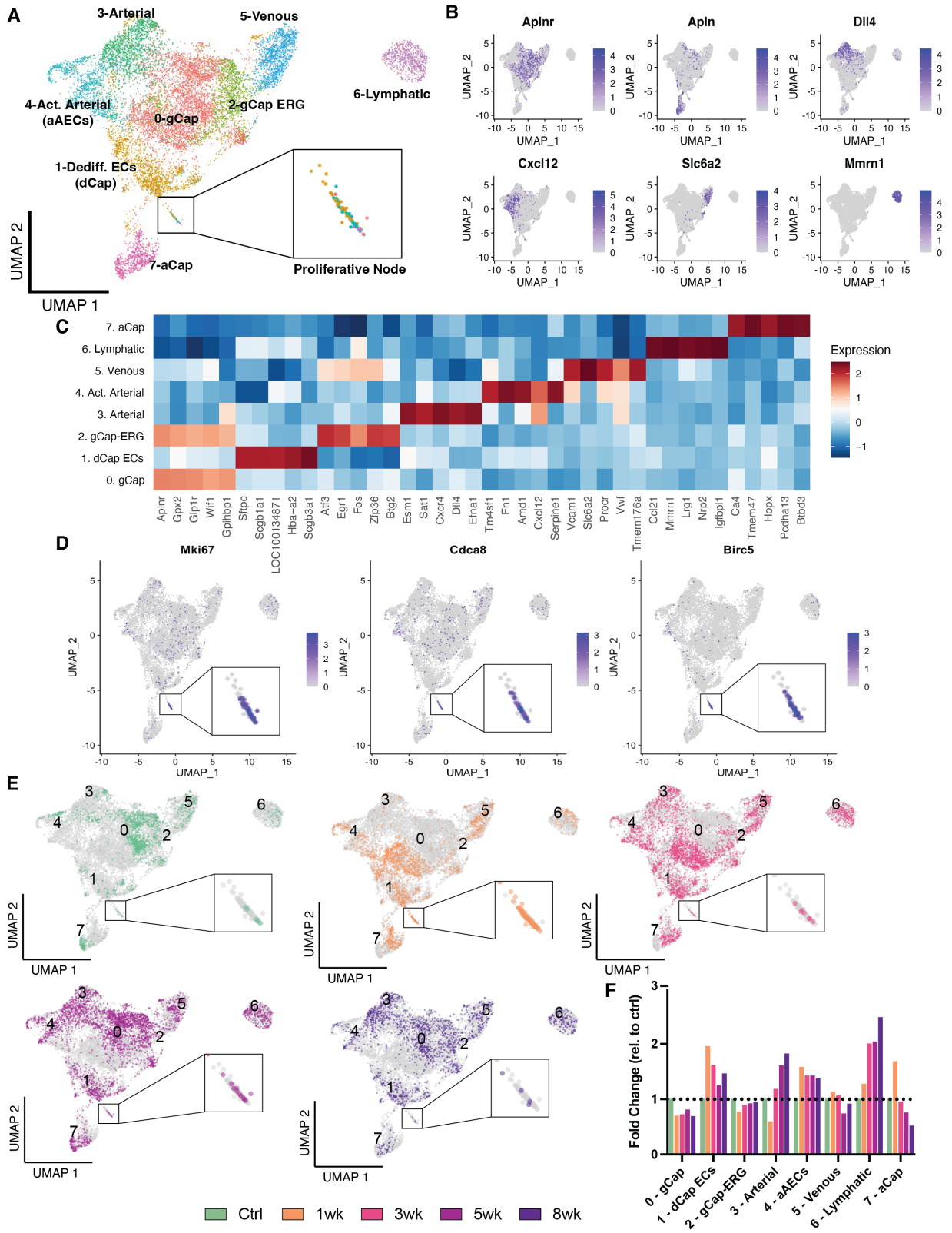

Figure 3: Endothelial cells during PAH progression. A) Uniform manifold approximation and projection (UMAP) representation of all endothelial populations at all sampled timepoints

showing 8 distinct cell types. B) Feature plots highlighting genes commonly associated with distinct endothelial cell (EC) populations including: *Aplnr* (gCap), *Apln* (aCap), *Dll4* (arterial), *Cxcl12* (activated arterial [aAECs] and arterial), *Slc6a2* (venous), and *Mmrn1* (lymphatic). C) Heatmap showing the top 5 differentially expressed genes distinguishing each EC population. D) Feature plots showing the distribution of proliferative genes largely within a proliferative node highlighted and enlarged in the box. E) UMAP of cells present at each timepoint demonstrating greater transcriptomic shifts in most populations at 1-, 3- weeks compared to control, 5-, and 8-weeks. F) Fold change of each EC populations relative to control levels demonstrating early and persistent increases in aAECs and dedifferentiated gCap (dCap) ECs, whereas as arterial and lymphatic ECs showed a later increase.

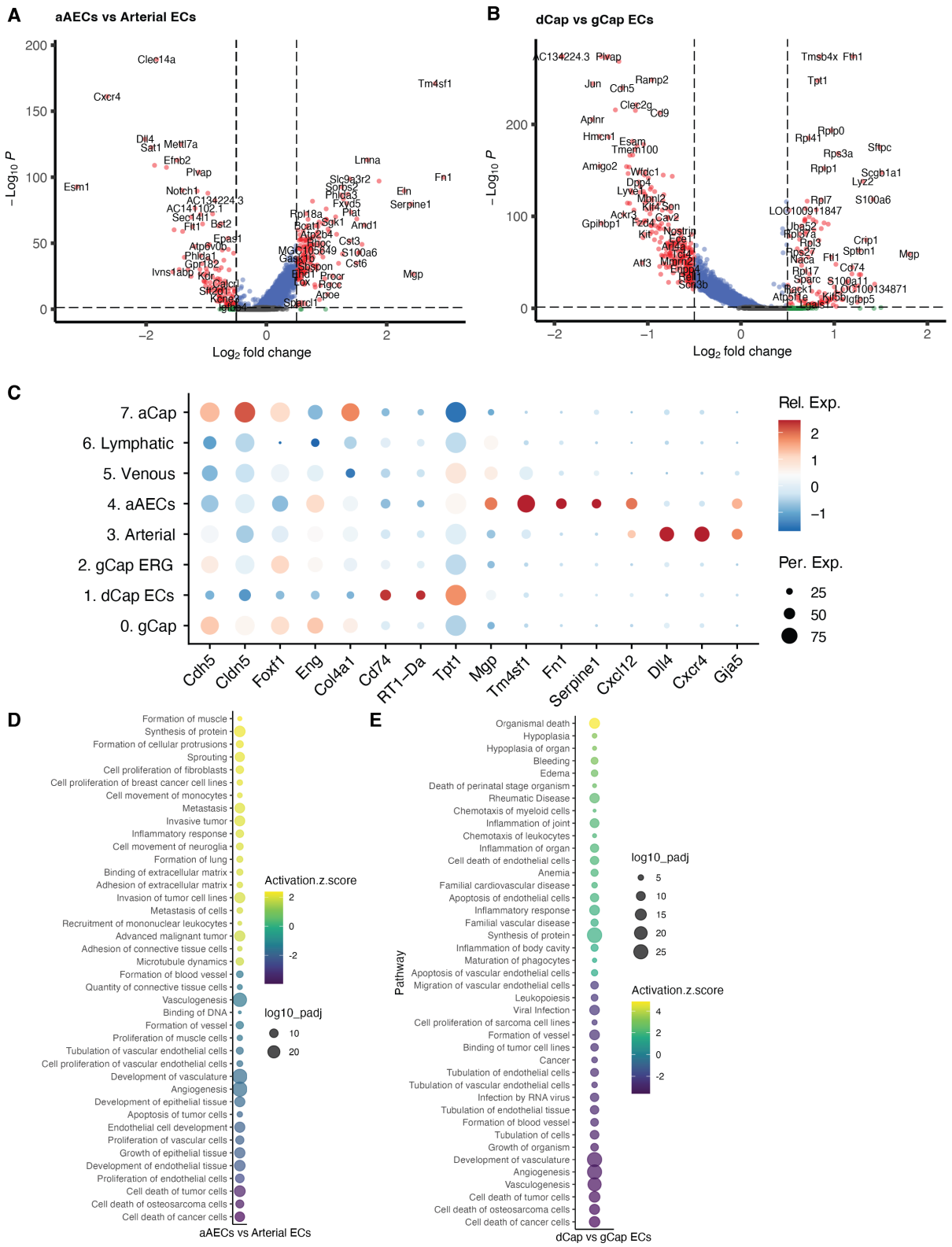

Figure 4: Transcriptomic signature of activated arterial and dedifferentiated EC populations. Volcano plots for all genes showing differential expression between aAECs/arterial ECs (A), and

dCap ECs/gCap ECs (B) across all timepoints. C) Dot plot showing relative expression of genes of interest. dCap ECs exhibited the lowest expression of *Cldn5* and *Cdh5*, and the highest expression of *Cd74*, *RT1Da* and *Tpt1*, whereas aAECs showed the highest expression of *Tm4sf1*, among other genes associated with an activated phenotype. D) Ingenunity pathway analysis (IPA) comparing aAECs and arterial ECs showing the top 15 significantly up and down regulated gene sets. aAECs exhibited upregulation of gene sets associated with cancer cell proliferation, migration and invasion, with downregulation of gene sets for angiogenesis, vascular development and apoptosis. E) IPA comparing dCap and gCap ECs showing the top 15 significantly up and down regulated gene sets, demonstrating upregulation of gene sets associated with inflammation and apoptosis, and downregulation of gene sets associated with vascular development and angiogenesis.

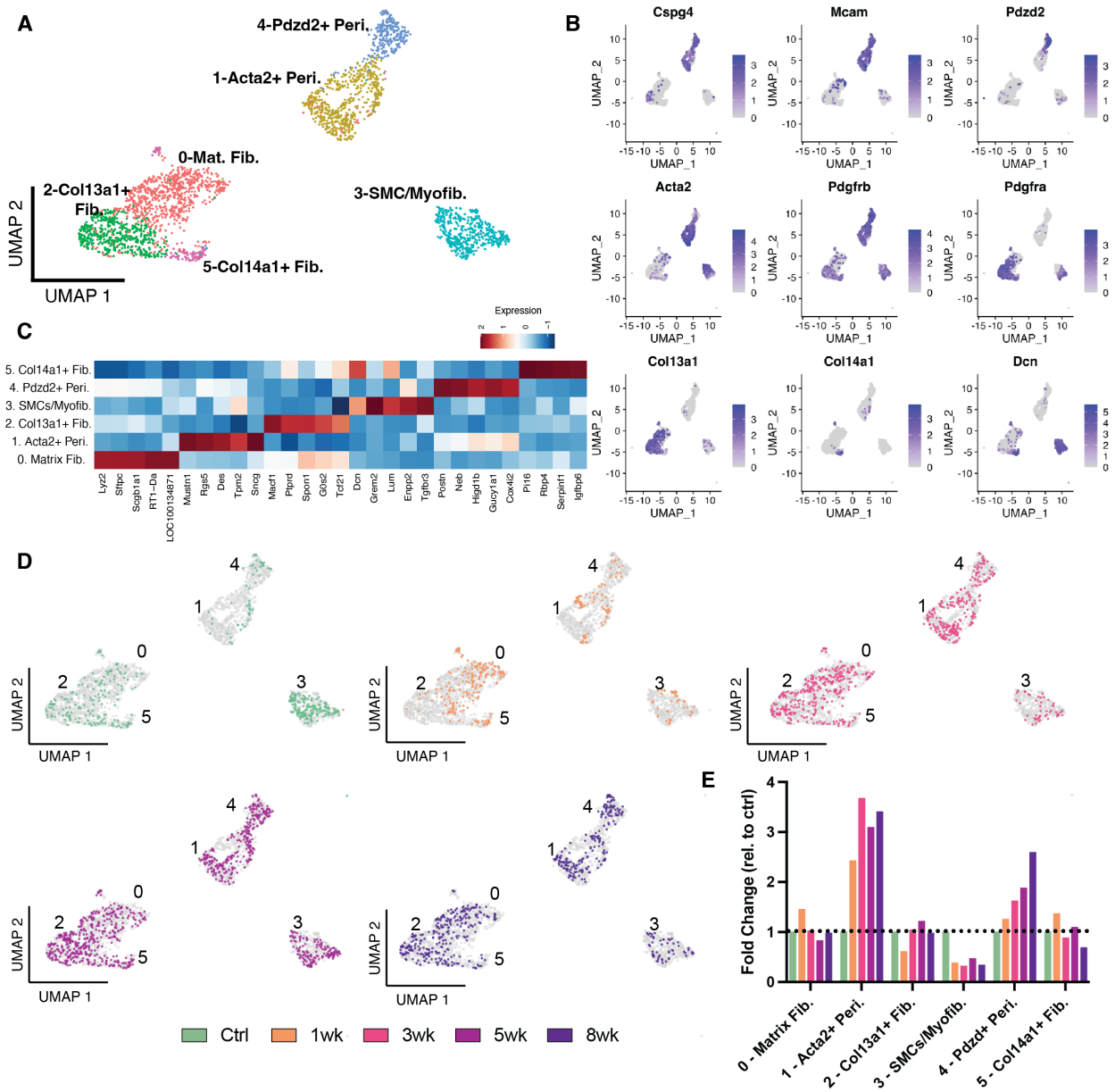

Figure 5: Stromal cells during PAH progression. A) Uniform manifold approximation and projection (UMAP) representation of all stromal populations at all sampled timepoints showing 6 distinct cell types. B) Feature plots highlighting genes typical of distinct stromal populations including: pericytes (*Cspg4, Mcam, Pdgfrb*); fibroblasts (*Pdgfra, Col13a1, Col14a1*); smooth muscle cells (*Acta2*); and myofibroblasts (*Dcn*). C) The top 5 differentially expressed genes distinguishing each stromal population. D) UMAP portraying the presence of cells at each timepoint, showing an expansion of both pericyte populations at 3-weeks. E) Fold change of each stromal population relative to control demonstrating proportional increases in both the classical and the smooth muscle (SM)-like *Acta2*+ pericyte populations during PAH progression.

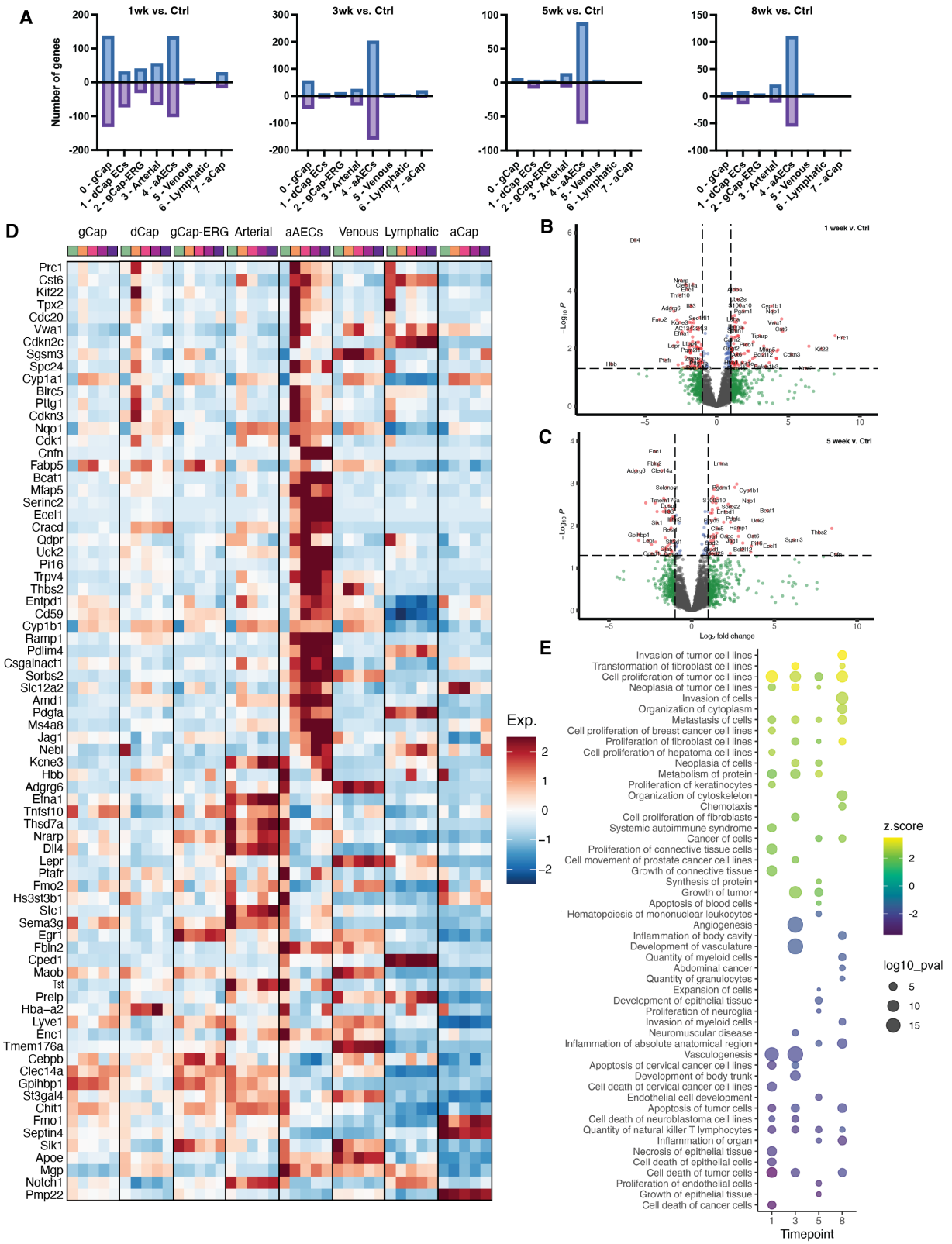

Figure 6: Endothelial transcriptomic changes during onset and progression of PAH. A) Differentially expressed genes (DEGs) in the various EC populations compared with control

(healthy) at 1, 3, 5 and 8 weeks post SU. A large number of DEGs was seen in many EC clusters at 1-week, but only aAECs displayed a persistently high number of DEGs at later timepoints. Volcano plots show all DEGs for aAECs at 1-week vs control (B) and 5-weeks vs control (C). D) heatmap of the top 20 up and downregulated DEGs for aAECs at each timepoint relative to control reveals three transcriptional patterns: (i) peak increase at 1-week with partial normalization at later timepoints (ii) persistent elevation at all timepoints, and (iii) transient or sustained reduction in gene expression. E) Ingenuity pathway analysis (IPA) for aAECs comparing each timepoint to the control cells, showing the top up or downregulated disease or biofunction gene sets by timepoint.

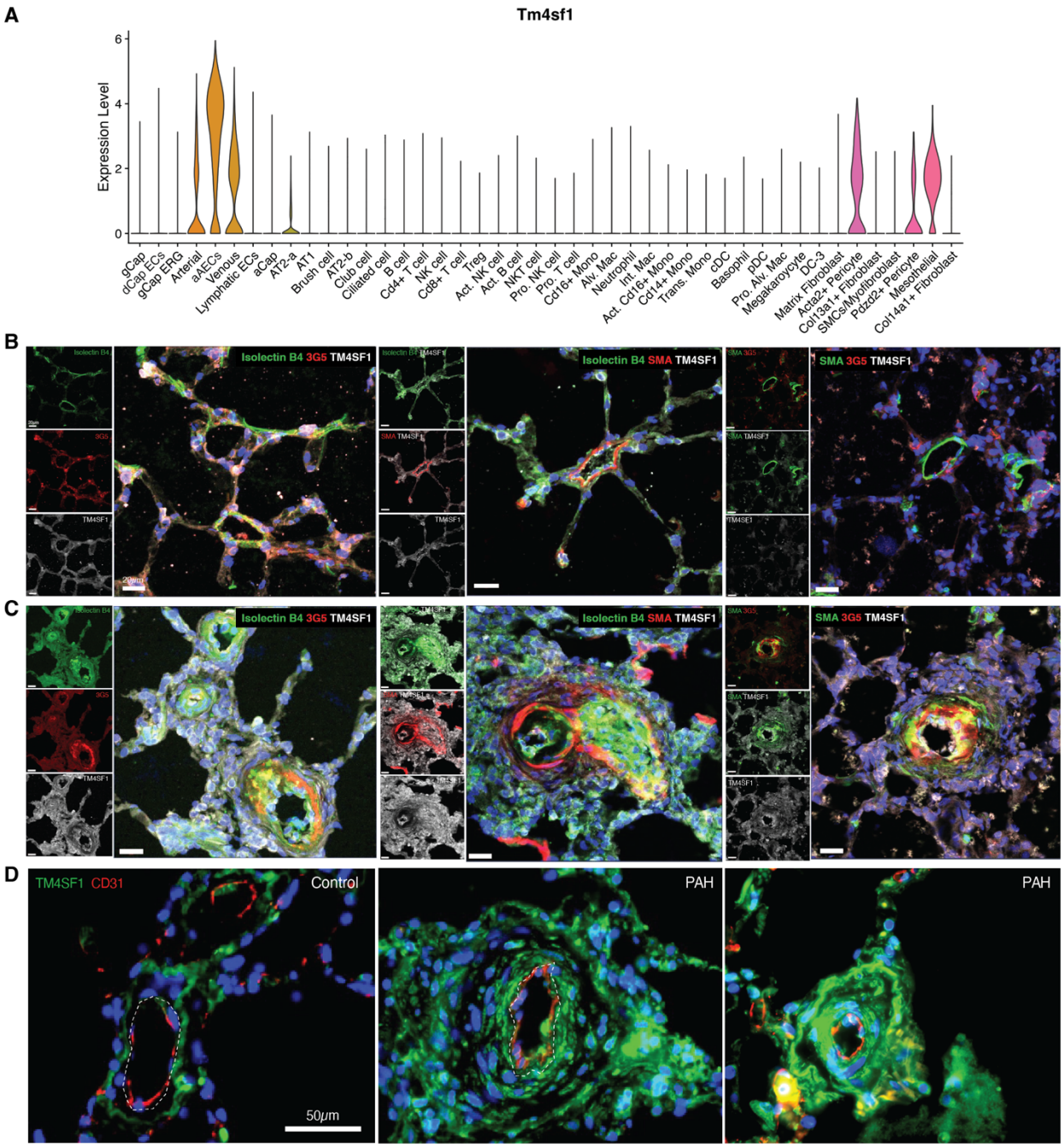

Figure 7: Tm4sf1 expression and vascular remodeling in PAH. A) Violin plots showing *Tm4sf1* expression mainly in arterial and venous ECs, pericytes and mesothelial cells, with the aAECs exhibiting the highest expression. Immunofluorescent staining for TM4SF1 with co-staining for ECs (isolectin B4), pericytes (3G5), or smooth muscle cells (SMA) in (B) healthy lungs (control) or (C) at 5-weeks in the SU/CH model. D) Immunofluorescent staining of lung sections from control subjects and PAH patients for TM4SF1, co-stained for ECs (CD31), demonstrating abundant TM4SF1 expression in PAH associated remodeled arteries relative to controls. Scale bar represents 20μm (B, C) and 50μm (D).

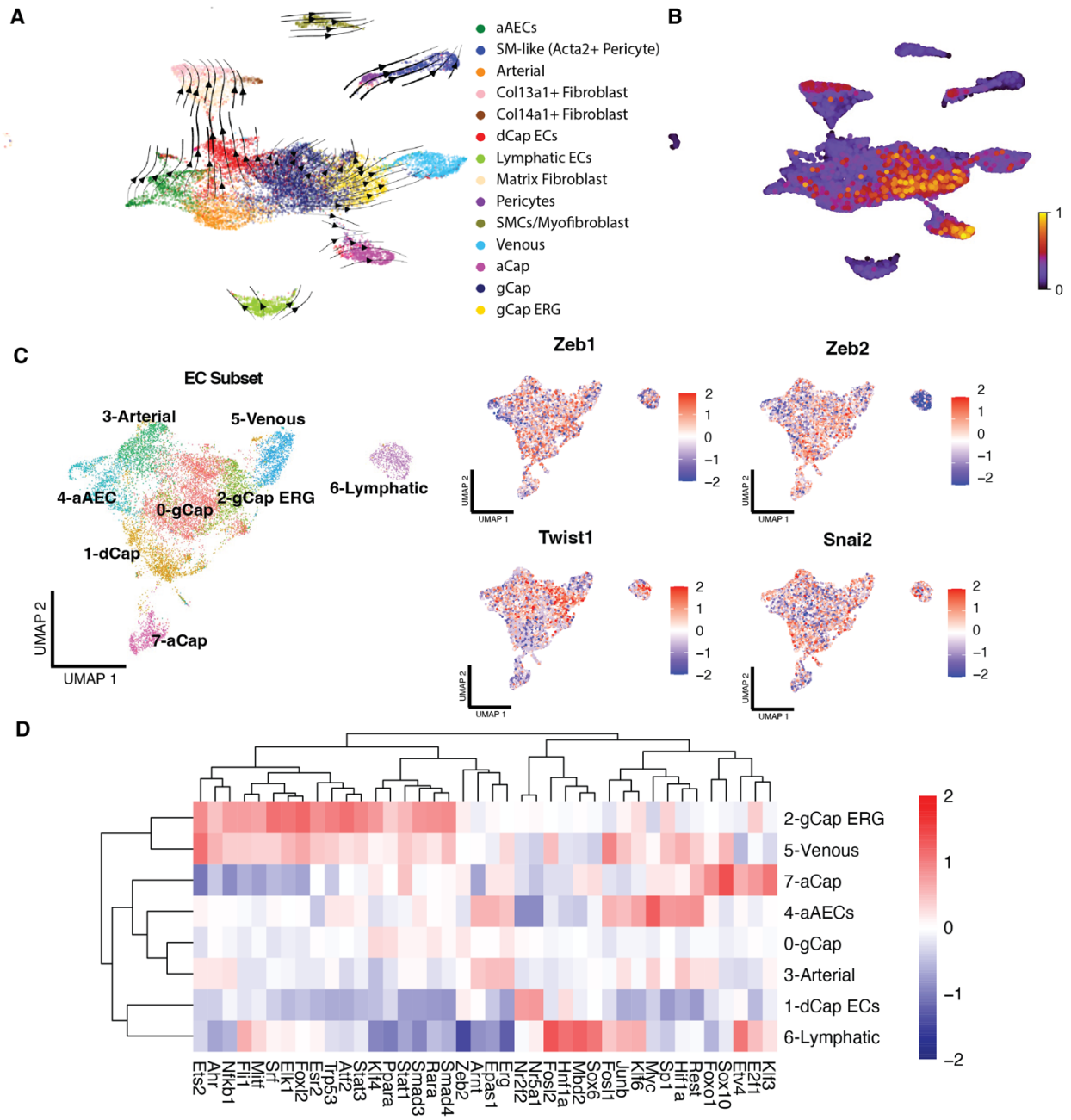

Figure 8: RNA velocity analysis of vascular cells and predicted EC transcription factor activity. A) RNA velocity analysis using scVelo's dynamical model and CellRank demonstrated strong vectors toward fibroblast populations mainly originating in the dCap EC cluster. Additionally, RNA velocity vectors suggest classical pericytes are transitioning towards the SM-like (*Acta2*+) pericyte fate. B) Latent time analysis provides an estimate of cellular differentiation, suggesting that gCap, aCap, fibroblasts, and classical pericytes represent a more terminal differentiated state, while aAECs, dCap and SM-like pericytes represent a more progenitor state. C) UMAP showing the distribution of typical EMT related transcription factor (TF) activities in the EC populations. E) Top 40 variable TFs in the different EC populations.