## SUPPLEMENTARY FIGURES

Figure S1. Total lung cluster and endothelial subclusters of cohorts 1 and 2 before integration. *Left*, cohort 1. *Right*, cohort 2. (A) UMAP of total lung clusters. (B) Violin plots for endothelial specific markers VWF (blood endothelial cells), PROX1 (lymphatic endothelial cells) and CALCRL (Pan-endothelial) in total lung. A full panel of marker were produced (please see materials and methods), only a selection are shown in this figure. (C) UMAP of total endothelial sub-cluster. (E) Violin plots for endothelial specific markers VWF (blood and lymphatic) clusters. (D) Heatmap of top 10 differentially expressed genes by endothelial sub-cluster. (E) Violin plots for endothelial specific markers VWF (blood endothelial cells), PROX1 (lymphatic endothelial cells) and CALCRL (Panendothelial) in total endothelial. The resolution was determined using clustree R package (Materials and Methods), with 0.5 value used for this analysis and throughput the study to avoid over-clustering (data not shown). (UMAP) Uniform Manifold Approximation and Projection.

**Figure S2. Differentially expressed genes by cluster for all clusters in total lung.** (A) Violin plots for the genes which were identified as differentially expressed by cluster, grouped by cluster in whole lung. Screening these DEG (between clusters) enabled the visualisation of their expression across all clusters to confirm differential expression as shown by unsupervised analysis in Seurat. (B) Violin plots for nFeature\_RNA and nCount\_RNA in the total lung.

**Figure S3. Differentially expressed genes by cluster for blood endothelial cells.** Violin plots for the genes which were identified as differentially expressed by cluster, grouped by cluster in blood endothelial data set.

**Figure S4. Expression of genes identified as differential in our analysis compared to whole lung sample.** (A) Dotplot of differentially expressed genes between blood endothelial cell subclusters in whole lung sample. (B) Table detailing gene identities, location, and classification of BEC differentially expressed genes. (C) Diagram of intracellular location of BEC differentially expressed genes. Created using biorender. (D) Table detailing gene identities, location, and classification of LEC differentially expressed genes.

**Figure S5. Summary of 'true' differential genes by cluster.** Violin plots of 'true' differential genes per cluster, grouped by cluster in blood endothelial data set.

**Figure S6. Expression of markers proposed by Schupp et al., 2021**. Violin plots of gene markers for clusters proposed in Schupp et al., 2021 and Sauler et al., 2022 in blood endothelial data set.

**Figure S7. Key regulators for blood endothelial cells. (A)** Violin plots for known regulator genes in blood endothelial cell data set. **(B)** Dotplot of expression of selected EC-sub-type marker genes across 12 identified BEC sub-clusters. Genes were selected from differential genes identified from the literature with confirmation of their sub-cluster specificity using violin plots (Supplementary Figure 4). (i) Gillich et al., 2019 (ii) Vanlandewijck et al., 2019; (iii) Iso et al., 2013; (iv) Nukala et al., 2021; (v) Ochiya et al., 2014; (vi) Herwig et al., 2016. When applied to 12 ageing human lung BEC subpopulations, the expression analysis of preselected by us previously reported markers of lung capillary EC in human and murine tissues (Gillich et al., 2020), aerocytes (Gillich et al., 2020), classical regulators of arterial and venous differentiation (Adams, 2003; Swift and Weinstein, 2009; Vanlandewijck et al., 2019; Su et al., 2018; You et al., 2005; dela Paz et al., 2009) and large vessels (Iso et al., 2013; Nukala et al., 2021; Ochiya et al., 2014; Herwig et al., 2016).

Figure S8. Heterogeneity of cell cycle in blood endothelial clusters. (A) UMAP of clusters labelled by cell cycle state. (B) Stacked bar charts comparing the cell cycle states by cluster in donor and IPF blood endothelium. (C) Table of statistical difference in cell cycle proportions in fibrosis by cluster. Statistical analysis was performed using chi- square test. P > 0.05 was considered significant. Cell cycle analysis (accounting for a possible impact of proliferation on transcriptional profiles) of 12 identified BEC sub-clusters revealed that the majority of cells in 8 sub-clusters (0, 2, 3, 4, 5, 6, 7 and 9) were in S phase, in three (1, 10 and 11) - in G1 phase and in one (8) – in G2/M phase. (UMAP) Uniform Manifold Approximation and Projection.

**Figure S9. Expression profile of BEC sub-clusters**. (A) Violin plots of nFeature\_RNA *left*, and nCount\_RNA *right*, in BEC sub-clusters. (B) Violin plots of cell type marker genes in BEC and total lung dataset. (C) Violin plots of endothelial cell progenitor cell genes in BEC sub-clusters.

Figure S10. Gene signatures proposed by Travaglini et al., 2020. Dot plot of signatures from Travaglini et al., 2020 in blood endothelial data set. Art = Artery, Vn = Vein, aCAP = Aerocyte, gCAP = general capillary, Bro1/2 = Bronchial.

Figure S11. Signature specificity by sub-cluster. (A) Ridgeplots of average module scores for signatures across all sub-clusters of blood endothelial cells. (B) Correlation matrix of created signatures for blood endothelial cell sub-clusters.

**Figure S12. Differentially expressed genes of blood endothelial cell populations young and old donors.** (A) UMAP of old and young endothelial cells. (A-C) the same colour key is used for all three figures. (B) UMAP of young and old endothelial cells split by sample age. (C) *Left, table detailing number of cells per cluster by sample age. Right,* stacked bar chart of percentage of each cluster to total endothelium split by sample age. (D) Heatmap of top 10 differentially expressed genes by cluster.

Figure S13. Cell type-specific composition of young donor lung (n=3) to old donor lung (n=4). (A) Dot plots of genes signatures from previous analysis (aged donor vs. IPF) heatmap in aged donor and young donor. (B) Dot plot of key regulator expression in aged donor and young donor. For more details, please see materials and methods. (C) Ridgeplots of average module scores for signatures across all clusters in young lung. (D) Ridgeplots of average module scores for signatures across all clusters in old lung. Sig 1 = signature 1 etc. (E) table of average module scores for all 10 signatures. *Left*, young. *Right*, old. S1=Signature 1 etc. Coral intensity represents a positive score for identified clusters. (F) Correlation matrix for proposed signatures in young (*left*) and old (*right*) lung. For details of analysis please see materials and methods. (UMAP) Uniform Manifold Approximation and Projection.

**Figure S14. Differentially expressed genes per cluster between donor and fibrosis linked to IPF signalling.** (A) Violin plots for differentially expressed genes between donor and fibrosis for each sub-cluster. (B) IPF signalling pathways identified using IPA.

Figure S15. Profiles of expression of gene sets relevant to endothelial cell biology processes in blood endothelial cell sub-clusters in donors and IPF patients. Published gene set enrichment analysis (GSEA) libraries were utilised for the assessment of expression of marker genes associated with ten selected key/relevant to EC biology processes. Samples were assigned module scores using the Seurat function AddModualScores based on genes used on GSEA website. Please see Materials and Methods section for further details, including scoring and interpretation. Full gene lists are available in Supplementary Table 8. Key colour code for sub-clusters is the same as in Figures 2-4. (A-J) Histograms (ridgeplot) of Endothelial cells differentiation (n= 5), endothelial-mesenchymal transition (Endo-MT) (n= 12), senescence (n= 79), apoptosis (n= 161), proliferation (n= 54), migration (n= 175), angiogenesis (n= 48), inflammation (n= 567), vasodilation (n= 36) and permeability (n= 40) scores in all blood endothelial sub-clusters in donors (*donor*) and IPF (*fibrosis*). Number of genes in sets are indicated in brackets. Crucially, histograms show the distribution of a score across sub-cluster (thus reflecting the heterogeneity of gene expression in individual cells within this subpopulation) compared to the total cluster/population. (K) Table of average module scores per cluster. *Coral red* scale proportional to positive module score.

Figure S16. Cell subtype-specific composition of ageing human lung lymphatic endothelial cell cluster from donors and IPF patients from two independent cohorts. (A) UMAP representation of all cells and cell sub-clusters from original lymphatic endothelial cell (LEC) cluster (from all 18 pooled samples) in Figure 1. Clusters were labelled 0-4 according to their signature, which is presented as (B) heatmap of top 5 differentially expressed genes by cluster. Details about each cell sub-population can be found in Supplementary Figure 13-15. (C) Violin plots of expression of pan-endothelial and lymphatic/LEC markers. (D) Dot plot of expression of genes used in annotation of identified lymphatic EC sub-clusters. (E) Pie charts detailing the proportion of cells per cluster in cell cycle stages. (F) Clusters were labelled according to their detected gene markers and annotated according to published literature (see main text) and signatures were created based on differentially expressed genes. (G) Ridgeplots of developed signatures across 5 clusters. (H) Table of module scores for proposed signatures by cluster. Coral represents a positivity for proposed signature in particular cluster, grey represents a cluster with relative negativity for all signatures. (I) Pseudotime lineage analysis of 5 LEC clusters. Colours of cluster are the same as in A. Identified DEG and the expression data of known lymphatic regulators were used to generate LEC sub-type-specific signatures, done in a similar fashion to BEC sub-clusters analysis. All signatures were tested for their specificity by comparing cell populations for their similarity to each sub-cluster. (UMAP) Uniform Manifold Approximation and Projection.

**Figure S17. Differentially expressed genes by cluster for lymphatics.** Violin plots for the genes which were identified as differentially expressed by cluster in lymphatic endothelial cell data.

Figure S18. Heterogeneity of cell cycle of lymphatic endothelial clusters. (A) UMAP of clusters labelled by cell cycle state. (B) Stacked bar charts comparing the cell cycle states by cluster in donor and IPF lymphatic endothelium. (C) Table of statistical difference in cell cycle proportions in fibrosis by cluster. Statistical analysis was performed using chi- square test. P > 0.05 was considered significant. (UMAP) Uniform Manifold Approximation and Projection.

**Figure S19. Proportions of lymphatic endothelial cell sub-clusters in health and IPF**. (A) UMAP of endothelial cells split by sample condition (B) *left* table detailing the number of cells per cluster split by condition. *Right* stacked bar chart of percentage contribution of each cluster to endothelial cell population. (C) UMAP of endothelial cell population split by sample condition and labelled by cohort. (D) Stacked bar chart of percentage contribution of each cluster to endothelial cell population by sample condition for cohort 1. (E) Stacked bar chart of percentage contribution of each cluster to endothelial cell population for cohort 2 (UMAP) Uniform Manifold Approximation and Projection.

Figure S20. Transcriptional heterogeneity of lymphatic endothelial cell clusters, linked to biological processes in health and disease. (A-I) Histograms (Ridgeplot) of Endothelial cells differentiation (n= 5), endothelial-mesenchymal transition (Endo-MT) (n= 12), senescence (n= 79), apoptosis (n= 161), proliferation (n= 54), migration (n= 175), angiogenesis (n= 48), inflammation (n= 567), vasodilation (n= 36) and permeability (n= 40) scores in all blood endothelial sub-clusters in donors (*donor*) and IPF (*fibrosis*). (K) Table detailing the p values for difference in module scores between donor and fibrosis. Samples were assigned module scores using the Seurat function AddModualScores based on genes used on Gene Set enrichment website. More details can be found in methods section. Statistical analysis was in the form of Shapiro Wilcoxon test and Mann-Whitney U test.

## SUPPLEMENTARY TABLES

Supplementary Table 1. Differentially expressed genes between clusters in total lung.

**Supplementary Table 2**. Caparison of identified total lung clusters in the present study to 6 recent single cell RNA sequencing studies.

**Supplementary Table 3.** Quantification of cell numbers and percentage of each cluster separated by cohort and sample condition of whole lung.

**Supplementary Table 4.** Correlation matrix of differentially expressed gene signatures from heatmap of blood endothelial cells.

**Supplementary Table 5.** Quantification of the mean number of cells in each sub-cluster of blood endothelial cells per sample split by cohort and condition.

**Supplementary Table 6.** Quantification of cell numbers and percentage of each cluster separated by cohort and sample condition of blood endothelial cells.

**Supplementary Table 7.** Differentially expressed genes identified between donor and fibrosis for each cluster, which were later used for IPA.

Supplementary Table 8. Gene libraries used for 'scoring'.

**Supplementary Table 9.** Summary table of cell numbers and percentage of each cluster separated by cohort and sample condition of lymphatic endothelial cells.