Transcriptional characterisation of human lung cells identifies novel mesenchymal lineage markers

Soula Danopoulos[1], Soumyaroop Bhattacharya [2], Thomas J. Mariani[2] and Denise Al Alam [1]

Affiliations: [1]Saban Research Institute, Children's Hospital Los Angeles, Los Angeles, CA, USA. [2]Pediatric Molecular and Personalised Medicine Program, University of Rochester, Rochester, NY, USA.

Correspondence: Denise Al Alam, Saban Research Institute, Children's Hospital Los Angeles, 4650 W. Sunset Boulevard, Los Angeles, CA 90027, USA. E-mail: dalalam@chla.usc.edu

ABSTRACT
Rationale: The lung mesenchyme gives rise to multiple distinct lineages of cells in the mature respiratory system, including smooth muscle cells of the airway and vasculature. However, a thorough understanding of the specification and mesenchymal cell diversity in the human lung is lacking.
Methods: We completed single-cell RNA sequencing analysis of fetal human lung tissues. Canonical correlation analysis, clustering, cluster marker gene identification and t-distributed stochastic neighbour embedding representation was performed in Seurat. Cell populations were annotated using ToppFun. Immunohistochemistry and *in situ* hybridisation were used to validate spatiotemporal gene expression patterns for key marker genes.
Results: We identified molecularly distinct populations representing "committed" fetal human lung endothelial cells, pericytes and smooth muscle cells. Early endothelial lineages expressed "classic" endothelial cell markers (platelet endothelial cell adhesion molecule/CD31 and claudin 5), while pericytes expressed platelet-derived growth factor receptor-β, Thy-1 membrane glycoprotein and basement membrane molecules (collagen IV, laminin and proteoglycans). We observed a large population of "nonspecific" human lung mesenchymal progenitor cells characterised by expression of collagen I and multiple elastin fibre genes (*ELN*, *MFAP2* and *FBN1*). We closely characterised the diversity of mesenchymal lineages defined by α_2-smooth muscle actin (*ACTA2*) expression. Two cell populations, with the highest levels of *ACTA2* transcriptional activity, expressed unique sets of markers associated with airway or vascular smooth muscle cells. Spatiotemporal analysis of these marker genes confirmed early and persistent spatial specification of airway (*HHIP*, *MYLK* and *IGF1*) and vascular (*NTRK3* and *MEF2C*) smooth muscle cells in the developing human lung.
Conclusion: Our data suggest that specification of distinct airway and vascular smooth muscle cell phenotypes is established early in development and can be identified using the markers we provide.

Introduction

The lung is a complex organ comprised of many cell types that differentiate to acquire a specific fate during the course of development. The interactions between these various cell types (*e.g.* epithelial, endothelial, *etc.*) are essential for the development of adequate respiratory function, including gas exchange structures, conducting airways and vasculature. Epithelial cell lineages in lung development and disease have been extensively studied in multiple mammalian species, but most of the data are from rodent models [1–6]. Fewer studies have focused on nonepithelial lineages [7–9], such that a thorough understanding of the specification and inter-relationships among the diverse mesenchymally derived cells in the lung is lacking. This is particularly the case for the developing human lung. A recent study assessed the distinct mesenchymal lineages in mouse lung and determined that there are numerous functional pathways, such as fibroblast growth factor signalling, that signal to the epithelium to promote epithelial homeostasis [10]. However, other studies, including our own, have demonstrated that cellular and molecular processes during lung development differ between mice and humans [11–13], including in the context of fibroblast growth factor signalling [14].

The lung mesenchyme includes multiple distinct cell lineages in the mature respiratory system, including airway smooth muscle cells (ASMCs) and vascular smooth muscle cells (VSMCs), endothelial cells, pericytes, lipofibroblasts, and parenchymal fibroblasts. An array of cell-specific markers demarcates some of these cell lineages, such as platelet endothelial cell adhesion molecule/CD31 (*PECAM1*) for endothelial cells, platelet-derived growth factor receptor-β (*PDGFRB*) for pericytes and α_2-smooth muscle actin (*ACTA2*) in pan-smooth muscle cells. However, markers to identify each cell type remain elusive and nonspecific for the majority, especially in the case of fibroblasts and VSMCs *versus* ASMCs. To date, there are no definitive markers allowing us to distinguish pulmonary ASMCs from VSMCs in the lung. Several pathological lung conditions are associated with either or both cell types. For example, pulmonary arterial hypertension is characterised by hyperproliferation and remodelling of VSMCs [15], reverting from a differentiated to a de-differentiated phenotype [16], whereas asthma is a disease in which patients have compromised breathing due to narrowing and remodelling of the airways, partially as a result of hyperproliferative and/or dysfunctional ASMCs [17]. In this study, we aimed to characterise the different mesenchymal cell lineages in the human developing lung in general and the ACTA2 expressing cells in particular.

Methods

Human fetal tissue collection

De-identified human fetal lung samples were obtained under Institutional Review Board approval from both the University of Southern California (Los Angeles, CA, USA) and the Children's Hospital Los Angeles (Los Angeles, CA, USA) (USC-HS-13-0399 and CHLA-14-2211).

Single-cell suspension preparation

Fetal human lung tissues were digested to single-cell suspensions according to the LungMAP protocol with minor adjustments, eliminating elastase and reducing the incubation time to 40 min [18]. Isolated cells were resuspended in freezing media (90% fetal bovine serum/10% dimethyl sulfoxide) at a concentration $\leqslant 60 \times 10^6$ cells·mL^{-1} and frozen until use.

Single-cell sequencing

Cell capture and library production was performed on the Chromium system (10X Genomics, Pleasanton, CA, USA). Sequencing was performed on a HiSeq4000 (Illumina, San Diego, CA, USA), with read alignment to Genome Reference Consortium Human Build 3. All single-cell sequencing data analysis was performed using Seurat version 2.1 (https://CRAN.R-project.org/package=Seurat). Filtered data were log transformed, scaled, integrated using canonical correlation analysis, clustered and represented by t-distributed stochastic neighbour embedding (t-SNE). Differential expression was defined using a parametric Wilcoxon rank sum test at a significance level of p<0.05. Pathway analysis and cell type association was performed using the ToppGene Functional Annotation tool (ToppFun) (https://toppgene.cchmc.org). For focused analysis of smooth muscle-like cells, we analysed only those cells with evidence for ACTA2 expression (raw counts >2; n=781 cells), and applied the aforementioned steps to identify subclusters and their associated cell types.

In situ hybridisation

Fluorescence *in situ* hybridisation (FISH) was performed on formalin-fixed, paraffin-embedded native human fetal lungs, sectioned at 5 µm, using RNAscope technology (Advanced Cell Diagnostics, Newark, CA, USA) following the manufacturer's protocol and as previously described [14].

Immunofluorescent staining

Immunohistochemistry was performed on formalin-fixed, paraffin-embedded native human fetal lungs at different gestational stages. Specimens were processed as previously described [19]. Briefly, embedded lungs were sectioned at 5 μm, rehydrated and boiled in a sodium citrate antigen retrieval solution for 12 min. Sections were incubated with primary antibodies (supplementary table S1) overnight at 4°C and then stained with appropriate fluorochrome-conjugated secondary antibodies.

Results

Unsupervised assessment of cell types in human embryonic lung

To characterise cellular heterogeneity in the developing human fetal lung, we performed single-cell RNA sequencing of protease-dissociated cells from 11.5 or 18.5 weeks gestation lung samples (one sample of each stage). We analysed data on a total of 3237 cells (1364 cells at 11.5 weeks and 1873 cells at 18.5 weeks), with an average detection of 2000–3000 genes per cell (supplementary figure S1). Canonical correlation analysis, implemented in Seurat, was used for data integration across subjects. This data object was used for analysis and visualisation by t-SNE. We identified 10 clusters of cells, along with corresponding cluster marker genes (figure 1a). Functional enrichment analysis defined nine of these clusters as identifiable lung cell populations, with one cluster remaining ambiguous (figure 1a and supplementary table S2). The majority of the cells (>85%) appeared to be of mesenchymal origin including a large population of human lung mesenchymal progenitor (HLMP) cells, which were annotated as matrix fibroblasts (n=1678; cluster 0) based upon similarity with murine cells. Our data also identified more recognisable mesenchymal populations including endothelial cells (n=388; cluster 1) and pericytes (n=300; cluster 3). A small fraction of cells (<2%) were identified as epithelial (cluster 8). We also identified a cluster of cells (n=302) with molecular attributes of both mesenchymal and epithelial origin, which we have defined as myoepithelial (cluster 2). Interestingly, we observed multiple immune cell populations including mixed lymphocytes (n=122; cluster 5), myeloid cells (n=79; cluster 7) and T-cells (n=53; cluster 9) in these tissues from early stages of lung development.

Assignment of cluster identities was further validated by examining the expression of marker genes associated with major cell types (figures 1b and 2). Many cluster markers included signature genes expected of the suggested cell type. For example, for matrix fibroblasts (cluster 0), there was high relative expression of decorin (*DCN*) and α_1 type collagen I (*COL1A1*), whereas for pericytes (cluster 3), there was high relative expression of *THY1* and *PDGFRB*. Expression patterns and levels varied (figure 2) for some individual canonical cell lineage markers. Some markers were present in a number of different clusters, such as elastin (*ELN*), whereas others were highly restrictive, such as *PECAM1* (figures 1b and 2), which only appeared in the endothelial cell cluster.

Spatial validation of mesenchymal cell lineages

We chose to focus on mesenchymal cell clusters due to their high proportional representation in our data set and the relative lack of characterisation of these cells in general, particularly in the human lung. RNAscope FISH and immunofluorescent staining was used for validation and spatial assessment of cluster marker genes for each of the different mesenchymal cell clusters. We studied spatial expression patterns throughout different gestational stages of native human fetal lung (11, 14, 16 and 18 weeks; n=3 for each) spanning the pseudo-glandular and canalicular stages of development. For matrix fibroblasts (cluster 0), immunofluorescent staining of fibrillin 1 (FBN1), COL1A1 and DCN displayed generally comparable localisation patterns, all indicative of lung matrix (figure 3). FBN1 was primarily localised within the ASMCs, as well as in the VSMCs (figure 3a–d). Co-staining of COL1A1 and DCN demonstrated that the two co-localised throughout development across multiple of the identified populations including, but not limited to, ASMCs and VSMCs (figure 3e–h). For endothelial cells (cluster 1), validation of characteristic cell markers (platelet endothelial cell adhesion molecule/CD31 and claudin 5 (CLDN5)) was performed. CD31 immunofluorescent staining showed well-established and organised vascular and endothelial networks (figure 4a–d) as early as 11 weeks (figure 4a). CLDN5 mesenchymal staining specifically localised to endothelial cells, but CLDN5 was also detected in epithelial cells (figure 4e–h). Validation of pericyte markers (cluster 3) demonstrated these cells were specified early on in the developing human lung as evident by PDGFRB staining as early as 11 weeks (figure 5a–d). Staining for microfibrillar-associated protein 2 (MFAP2), a marker highly expressed in fibroblasts/stromal cells (cluster 4), demonstrated strong localisation around airway and vascular structures (figure 5e–h).

Discrimination of early human lung smooth muscle-like cell phenotypes

We were particularly interested in further distinguishing the large number of cells appearing to be of mesenchymal origin. We focused our efforts on populations that were defined as smooth muscle-like based upon the detection of ACTA2 expression (n=781). We analysed 3390 genes with highly variable
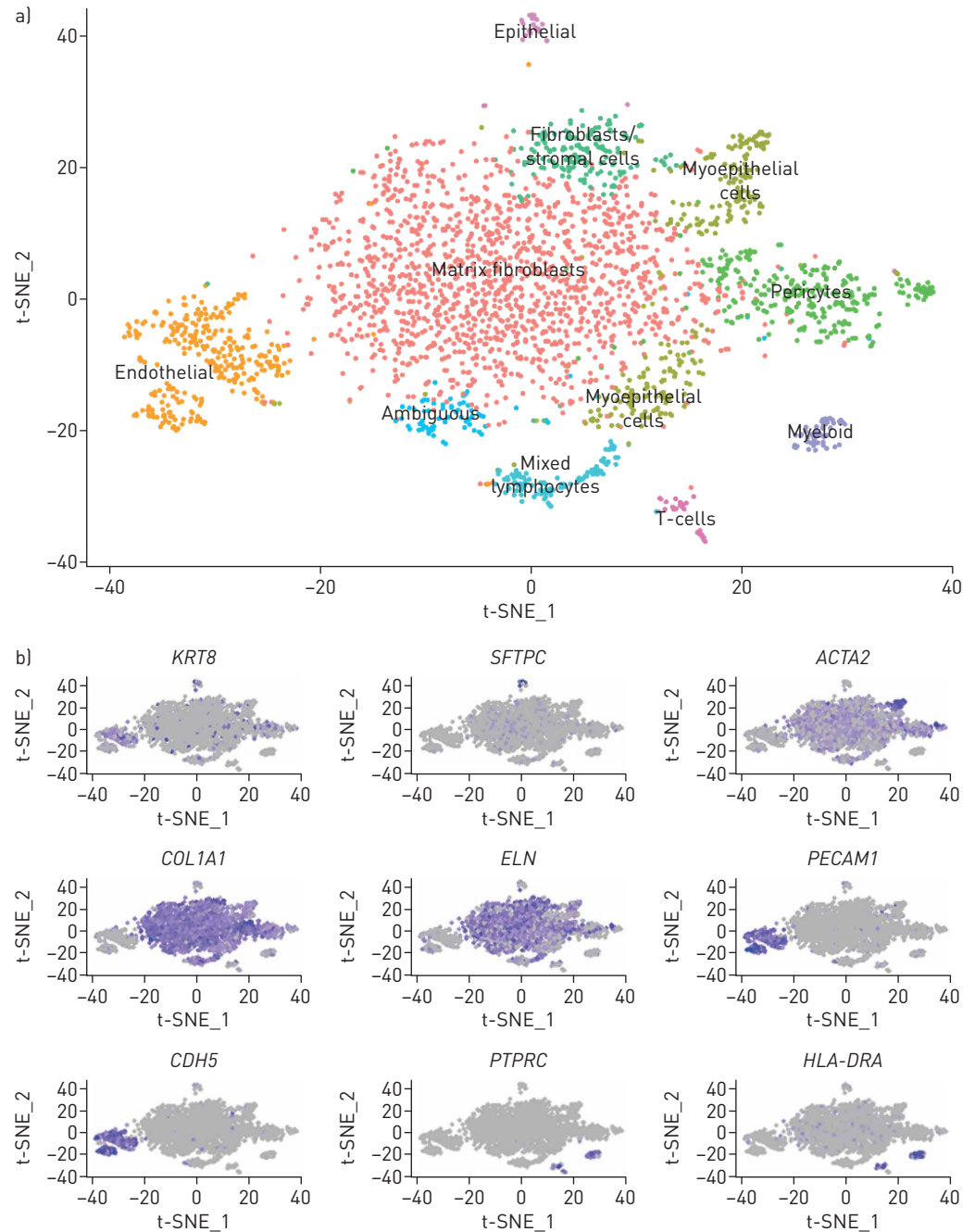
FIGURE 1 Single-cell RNA sequencing identifies cell populations in human fetal lung. t-SNE: t-distributed stochastic neighbour embedding. a) Unsupervised graph-based clustering of single-cell RNA sequencing data from early stages of lung development, visualised using t-SNE. Each point represents a single cell, and individual clusters are coloured and annotated based on cell type associations derived from ToppFun. b) Gene expression patterns for individual canonical cell lineage markers in fetal lung cell clusters, overlaid on t-SNE plots defining cell type clusters. Each point represents a single cell, with blue colour indicating expression level of the specified marker gene (darker shade is higher expression). Expression of some lineage markers is highly restricted (*e.g. SFTPC* to epithelial cells, *PECAM1* and *CDH5* to endothelial cells, and *PTPRC* and *HLA-DRA* to leukocytes), while other markers are more widely expressed (*KRT8, ACTA2, COL1A1* and *ELN*).

expression in the 781 cells with evidence of ACTA2 expression. Among the 10 subclusters identified for ACTA2$^+$ cells (figure 6a), a majority of them displayed a low to moderate level of ACTA2 expression (figure 6b). A majority of the 10 subclusters were annotated as various classes of identifiable mesenchymal cells including fibroblasts (n=112), myofibroblasts (n=264) and pericytes (n=31). Some of the subclusters were annotated as identifiable smooth muscle-like cells including vascular smooth muscle, bronchial smooth muscle, stromal cells of the trachea or fibroblasts of the aorta (figure 6d). Two subclusters clearly
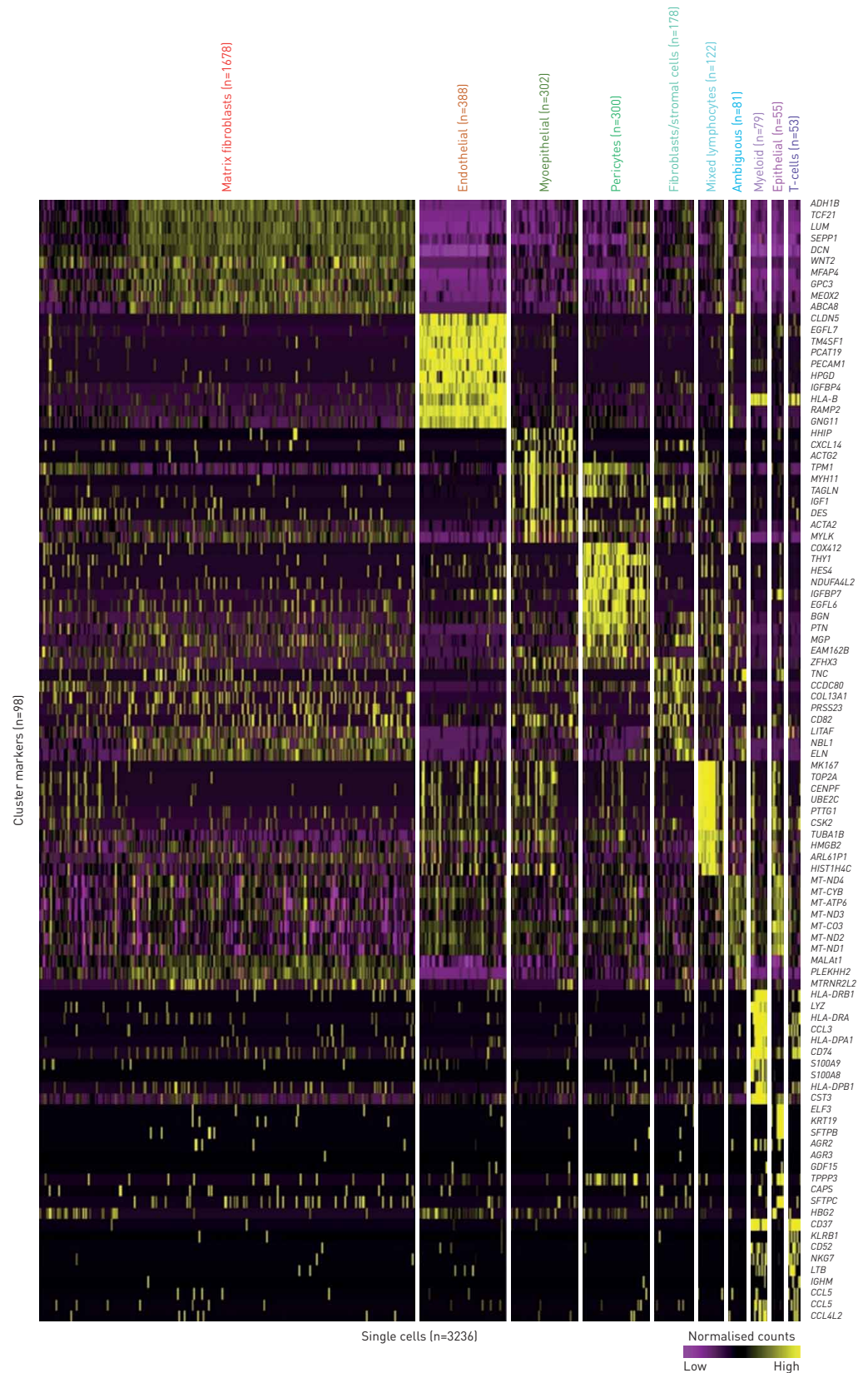
FIGURE 2 Heatmap indicating human fetal lung cell population markers. The heatmap displays gene expression patterns for marker genes (n=10 each; total of 98 due to two markers in more than one cluster) for each cell population cluster, based on differential expression testing. Individual genes are represented in rows and individual cells (n=3236) are represented in columns. Yellow indicates high relative gene expression and purple indicates low or no expression. Each cluster is labelled by its presumed cell type based upon annotation with ToppFun.
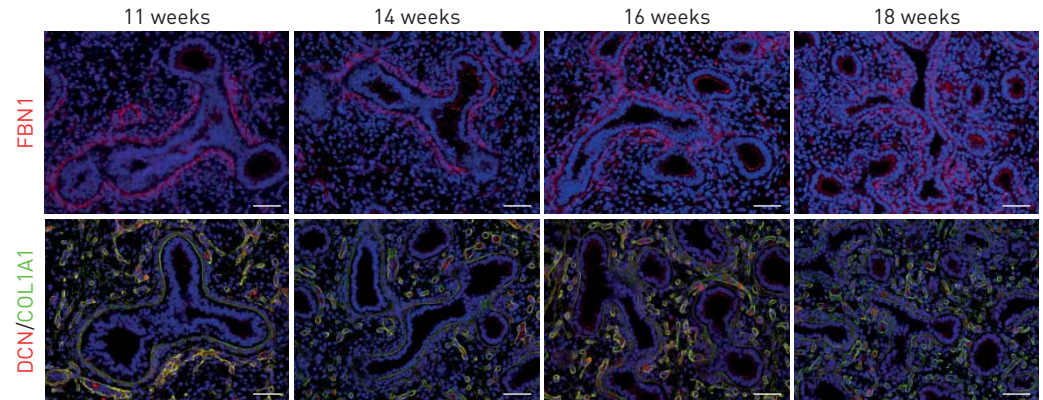
FIGURE 3 Spatial validation for matrix fibroblasts. a–d) Immunofluorescent staining of fibrillin 1 (FBN1, red) on a) 11, b) 14, c) 16 and d) 18 weeks fetal human lung demonstrates localisation within the airway and vascular smooth muscle cells. e–h) Immunofluorescent co-staining of decorin (DCN, red) and α_1 type collagen I (COL1A1; green) on e) 11, f) 14, g) 16 and h) 18 weeks fetal lung demonstrates that the two co-localise throughout development in multiple cell populations. Scale bar: 50 µm. n=3 for each gestational stage.

displayed the highest levels of ACTA2 expression (ACTA2^{hi}). These two ACTA2^{hi} subclusters had unique (as opposed to multiple) annotations for smooth muscle-like cells, with subcluster 0 markers annotated only as stromal cells of the trachea and subcluster 3 markers annotated only as vascular smooth muscle. Marker genes for these two subclusters are shown in figure 6c (and supplementary table S3) and demonstrate distinct transcriptional profiles for these two clusters.

Spatial validation of ASMC and VSMC markers

We examined the spatial expression patterns for ACTA2^{hi} subcluster 0 and subcluster 3 marker genes, to test the hypothesis that they represent markers for airway and vascular smooth muscle, respectively. RNAscope FISH was performed on human fetal lung tissue at 11, 14, 16 and 18 weeks gestation for selected subcluster markers (n=3 for each time-point). Subcluster 0 was characterised by high expression of fibroblast growth factor 18 (*FGF18*), insulin-like growth factor I (*IGF1*), myosin light chain kinase (*MYLK*) and Hedgehog-interacting protein (*HHIP*) (figure 6c). FISH of *HHIP*, *MYLK* and *ACTA2* clearly showed that at each gestational stage all three markers were coexpressed in the ASMCs (figure 7a–a′′′), with neither *HHIP* nor *MYLK* being expressed in the VSMCs. Similarly, *IGF1* and *FGF18* were coexpressed with *ACTA2* in ASMCs only but not in VSMCs (figure 7b–b′′′ and c–c′′′). High magnification panels of these markers with airway smooth muscle *ACTA2* seen in figure 7a′′′′, b′′′′ and c′′′′ and supplementary figure S4 highlight the extent of co-localisation. Subcluster 3 was characterised by high expression of Hairy/enhancer-of-split related with YRPW motif protein 2 (*HEY2*), neurotrophic receptor tyrosine kinase 3
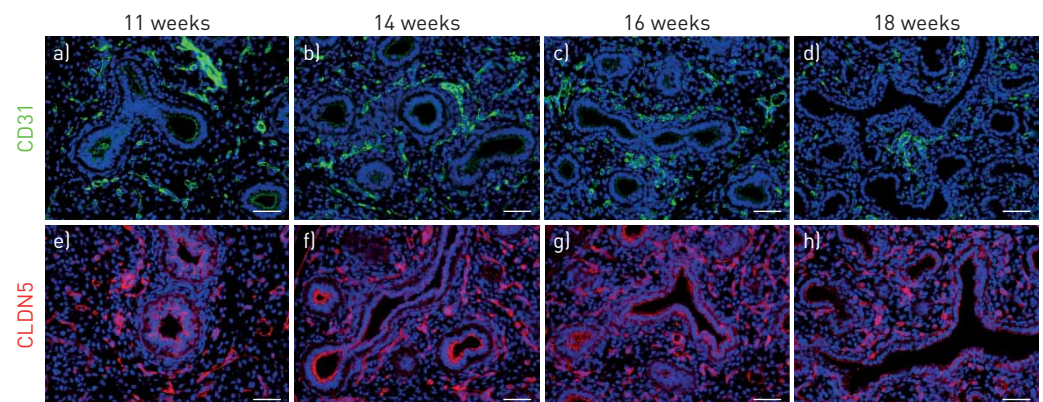

FIGURE 4 Spatial validation of endothelial cells. a–d) Immunofluorescent staining of platelet endothelial cell adhesion molecule (CD31; green) on a) 11, b) 14, c) 16 and d) 18 weeks fetal human lung localised specifically in the endothelial cells. e–h) Immunofluorescent staining of claudin 5 (CLDN5, red) on e) 11, f) 14, g) 16 and h) 18 weeks human fetal lung is primarily localised in the endothelial cells with some expression in the epithelial cells. Scale bar: 50 µm. n=3 for each gestational stage.
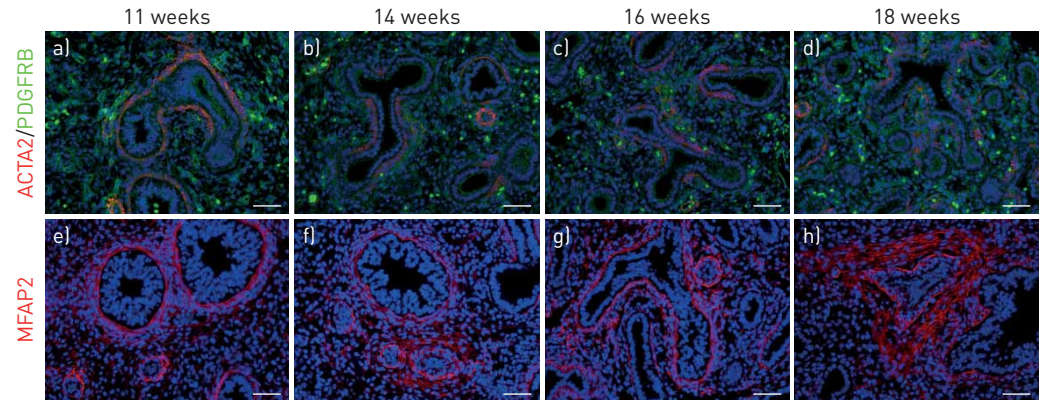
FIGURE 5 Spatial validation of pericytes and stromal cells. a–d) Immunofluorescent co-staining of α₂-smooth muscle actin (ACTA2; red) and platelet-derived growth factor receptor-β (PDGFRB; green) on a) 11, b) 14, c) 16 and d) 18 weeks native fetal human lung demonstrates the two are not coexpressed and the presence of pericytes from an early developmental stage. e–h) Immunofluorescent co-staining of microfibrillar-associated protein 2 (MFAP2; red) on e) 11, f) 14, g) 16 and h) 18 weeks native human fetal lung demonstrates strong localisation around airway and vascular structures. Scale bar: 50 μm. n=3 for each gestational stage.

(*NTRK3*) and myocyte enhancer factor 2C (*MEF2C*) (figure 6c). FISH of *NTRK3*/*MEF2C* and *ACTA2* clearly demonstrates that these two markers are both predominantly coexpressed in VSMCs with extremely low expression of *MEF2C* in ASMCs (figure 8a–a′′′). Similarly, *HEY2* is coexpressed with *ACTA2* in VSMCs but not ASMCs. *HEY2* is also detected at low levels in some airway epithelial cells (figure 8c–c′′′). The coexpression of these markers with vascular smooth muscle *ACTA2* is also observed in the high magnification panels (figure 8b–b′′′ and d–d′′′).

Discussion

We carried out the current study to expand our understanding of cellular phenotypes of the fetal human lung and their molecular attributes. Prior studies by many investigators have described similar information from the adult human lung in disease states (*e.g.* idiopathic pulmonary fibrosis *versus* controls) [20, 21] and during development of animal models, most notably from the mouse [22–24]. The LungMap consortium has begun to define various cell types in the developing mouse lung based on single-cell RNA sequencing [25]. As multiple recent studies have uncovered substantive differences in the expression of key molecules during mouse and human lung development, our studies aimed to provide a comprehensive data set with which to assess these similarities and differences on a global level. Additional analysis, and future work, will be necessary to further quantify the extent of the similarities. However, our data set already provides novel information that can be used to identify markers of discrete cell populations within the fetal human lung.

Our study has some limitations; mainly the fact that the single-cell RNA sequencing data presented are derived from cells recovered from only two samples, one each at 11.5 and 18.5 weeks. It is important to note, however, that statistical power for the analysis is provided by the fact that each population or cluster of cells was represented by dozens or hundreds of independent cell gene expression profiles. Moreover, the validation by *in situ* hybridization/immunofluorescence was performed on four different stages of lung development and three independent sample per stage, which all show consistent results. Critically, each cluster contained cells from both samples (supplementary table S2). Although the distribution of cells was not identical among samples, we were not able to determine whether these differences were biologically meaningful. Because of our interest in mesenchymal cells, and in an attempt to enrich these cells and recover as many live mesenchymal cells as possible, we shortened the dissociation time, which resulted in a reduction of the number of epithelial cells recovered. Therefore, our dissociation protocol is somewhat selective for mesenchymal cells, as we recovered a very low percentage of epithelial cells (<2%). Importantly, consistent with our data, recent work suggests that standard single-cell dissociation methods lead to recovery of lower than expected numbers of epithelial cells in paediatric human lung tissues [18]. Notably, the frequency increases with age and is maximal in adult tissues. However, this did not affect our ability to analyse the different mesenchymal populations in the human fetal lung.

From the data presented here, we can infer that specification of diverse populations of mesenchymal cells, including discrete fibroblast, pericyte and smooth muscle cell populations, initiates prior to 11 weeks gestation in humans. Conversely, our data indicate that the fetal human lung contains a large number of "nonspecific" HLMP cells. RNA sequencing analysis indicates that HLMP cells are characterised by the
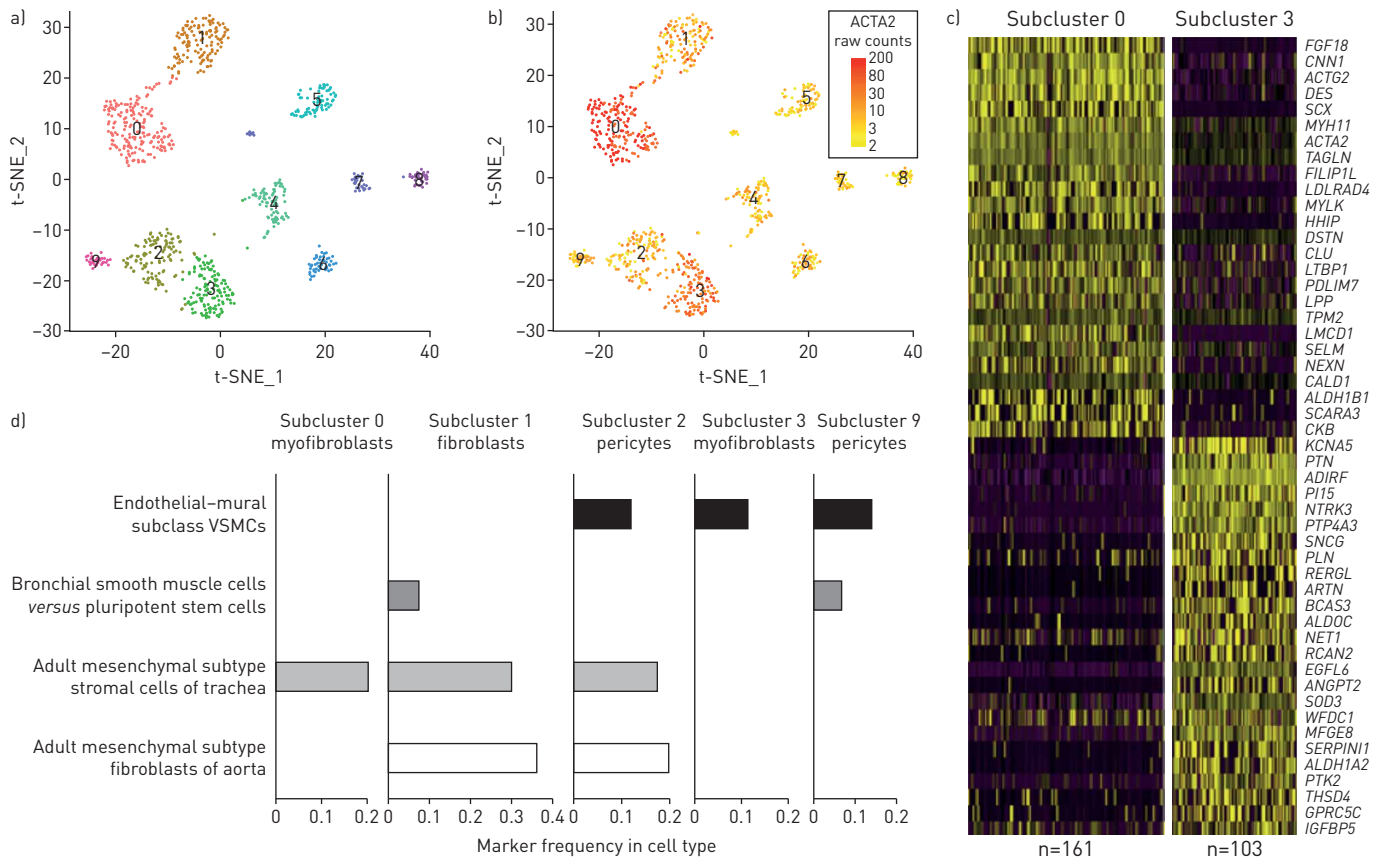
FIGURE 6 Heterogeneity of cells expressing α_2-smooth muscle actin (ACTA2). t-SNE: t-distributed stochastic neighbour embedding; VSMC: vascular smooth muscle cell. a, b) Unsupervised graph-based subclustering of single-cell RNA sequencing data of cells expressing ACTA2 (raw count >2; n=781) visualised using t-SNE. Each point represents a single cell, identified as a separate population by a) unsupervised clustering or b) coloured based on the expression level of ACTA2. Interestingly, cells with highest ACTA2 (ACTA2^{hi}) expression reside in one of two subclusters (0 and 3), whose marker genes identify them as myofibroblasts. c) Heatmap displaying gene expression of cluster marker genes (n=25 each) for the two ACTA2^{hi} cell populations (subclusters 0 and 3). Individual genes are represented in rows and individual cells (n=264) are represented in columns. Yellow indicates high relative gene expression and purple indicates low or no expression (see figure 2 for colour scale). d) In-depth cell type annotation analysis using marker genes for ACTA2 expressing populations: frequencies for marker genes being annotated to specific lung (LungMAP) cell types. Multiple ACTA2 subclusters have several smooth muscle/mesenchymal stromal cell-related annotations (subclusters 1, 2 and 9). However, the two ACTA2^{hi} subclusters have unique smooth muscle cell annotations, one associated with the trachea (subcluster 0) and one associated with the vasculature (subcluster 3).

expression of interstitial collagens and multiple elastic fibre genes, including tropoelastin (*ELN*). *ELN* expression in the lung is thought to be highly restricted in expression at later stages [26] and in rodents [27], whereas in the human fetal lung, it is abundantly expressed across a large number of cell types. Interestingly, HLMP cells appear to be highly similar to so-called matrix fibroblasts observed in the developing mouse lung [25].

Our data also demonstrate that immune cells derived from the haematopoietic system, including lymphoid and myeloid cells, occupy the fetal human lung at very early stages. Macrophages are known to play a role during lung development in rodents [28]. However, our data are the first to suggest that localisation of varied immune cells occurs during normal human fetal lung development. The identity of these immune cells is still poorly defined at this time, but should be investigated in the future to better understand their function and to determine their possible mechanistic role during lung development.

Myoepithelial cells are multipotent stem cells, present at the base of the submucosal glands, that have the potential to give rise to all cell types of the submucosal glands and surface epithelium in response to injury [29]. These cells are characterised by the expression of both epithelial and mesenchymal cell markers (P63, keratin 5 (*KRT5*) and ACTA2). In mouse lung single-cell analyses, myoepithelial cells were not detected [25], likely because submucosal glands are present only in the upper part of the trachea in the mouse lung, whereas in the human lung the submucosal glands extend from the trachea to the terminal bronchioles. To exclude low-quality cells, we filtered out cells having less than 500 unique molecular identifiers detected or with ⩾12.5% of genes being mitochondrial genes, which were excluded for being likely doublets (supplementary figure S3). We have not completely confirmed that this approach
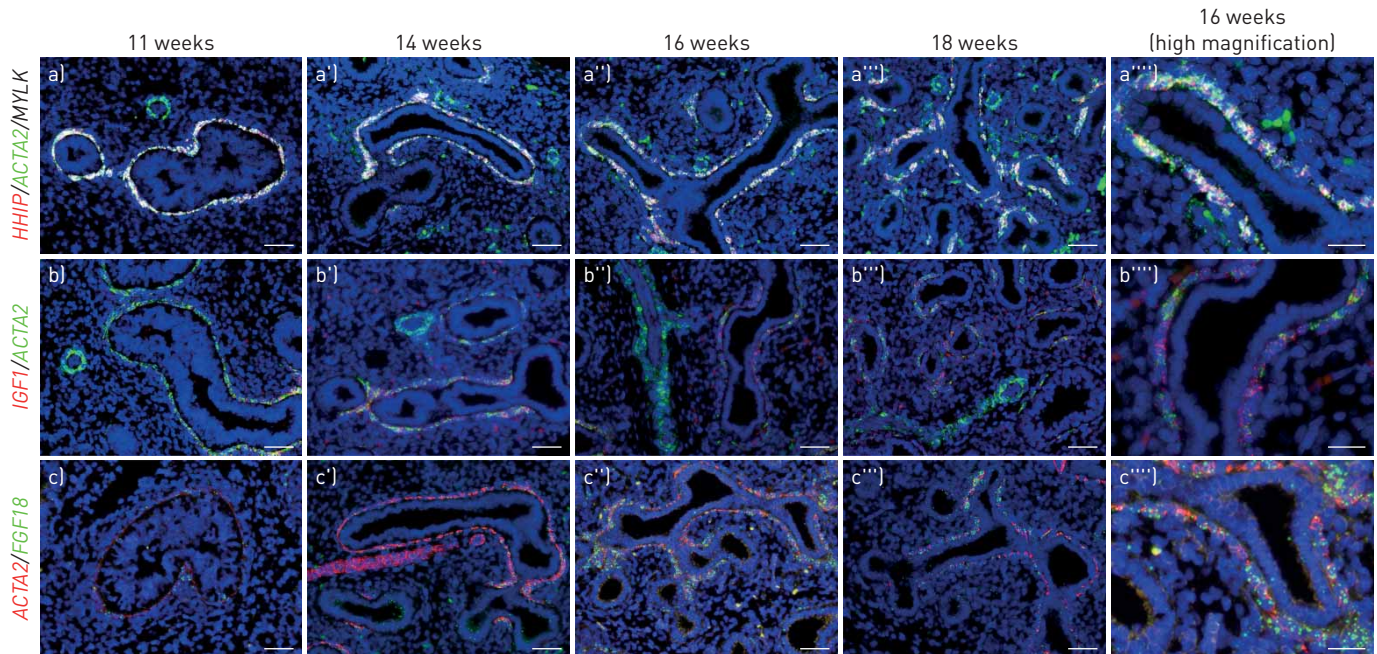
FIGURE 7 Airway smooth muscle cell (ASMC) spatial validation. FISH: fluorescence *in situ* hybridisation; VSMC: vascular smooth muscle cell. a–a′′′) FISH of Hedgehog-interacting protein (*HHIP*; red), myosin light chain kinase (*MYLK*; white) and α_2-smooth muscle actin (*ACTA2*; green) on fetal human lung sections at a) 11, a′) 14, a′′) 16 and a′′′) 18 weeks gestation demonstrates co-localisation around the ASMCs with no presence around the VSMCs. b–b′′′) FISH of insulin-like growth factor 1 (*IGF1*; red) in conjunction with *ACTA2* (green) on fetal human lung sections at different gestational stages demonstrates that IGF1 is solely localised in the ASMCs throughout development. c–c′′′) FISH of fibroblast growth factor 18 (*FGF18*; green) in conjunction with ACTA2 immunofluorescence (red) demonstrates that FGF18 is localised in the ASMCs and lower expression in the distal epithelium between 11 and 18 weeks of lung development but is not expressed in the VSMCs. Scale bar: 50 µm. High magnification panels of 16 weeks gestation confirm co-localisation of markers within the ASMCs (a′′′′, b′′′′ and c′′′′). High magnification panels for other gestations are provided in supplementary figure S4. Scale bar: 25 µm. n=3 for each gestational stage.

successfully removed doublets in our data set and cannot completely exclude the possibility that the myoepithelial cell cluster contains some doublets. We did explore the relative quality control metrics across individual clusters (supplementary figure S3), and clearly see that the myoepithelial cell cluster (cluster 2) is near the mean for the number of genes detected per cell, the number of unique molecular identifiers per cell and the percentage of mitochondrial genes per cell. These data suggest that the cells in this cluster are not of different quality to those in other clusters that represent individual recognised cell types. Importantly, the following features also support the biological, rather than technical, nature of this cluster: substantial proportions of cells from both samples were assigned to this cluster (90 and 212 cells, respectively; supplementary table S2), the total number of genes detected in this population is similar to cells of other defined phenotype/annotation and the number of cluster-specific marker genes is similar to cells of other defined phenotype/annotation. Therefore, the abundant presence of a population of myoepithelial cells in our analyses was interesting but not surprising. Their role and contribution to the development of the human lung is unknown and requires further investigation.

Mesenchymal cell types of the lung have been poorly defined relative to epithelial cell types. ASMCs and VSMCs share similar contractile functions, and both express *ACTA2*, but are located at distinct locations. A significant knowledge gap exists regarding the lack of specific and exclusive markers to differentiate them. Encouraged by the diversity of ACTA2$^+$ cells revealed by subclustering of fetal human lung RNA sequencing data, we tested populations for attributes of frank smooth muscle cells. Our ontology analysis suggested that the two subclusters with the highest level of *ACTA2* expression may represent ASMCs and VSMCs, respectively. FISH analysis confirmed that expression of at least a subset of markers differentially expressed in these two ACTA2^{hi} populations was spatially segregated and unique to either ASMCs (*e.g. MYLK* and *HHIP*) or VSMCs (*e.g. MEF2C, NTRK3* and *HEY2*). Among the marker genes for these two ACTA2^{hi} populations, which appear to be ASMCs and VSMCs, respectively, there are a number of genes that encode cell surface proteins (supplementary table S3). Validation of these markers in future studies will be necessary to determine if these markers alone, or in combination with others, can be utilised to isolate/sort these cell populations.

Interestingly, mutations in some markers identified in VSMCs, such as *NTRK3, HEY2* and *MEF2C*, are associated with congenital heart disease [30–33], whereas *HHIP*, identified in the ASMCs, has been associated in genome-wide association studies with chronic obstructive pulmonary disease (COPD) [34].
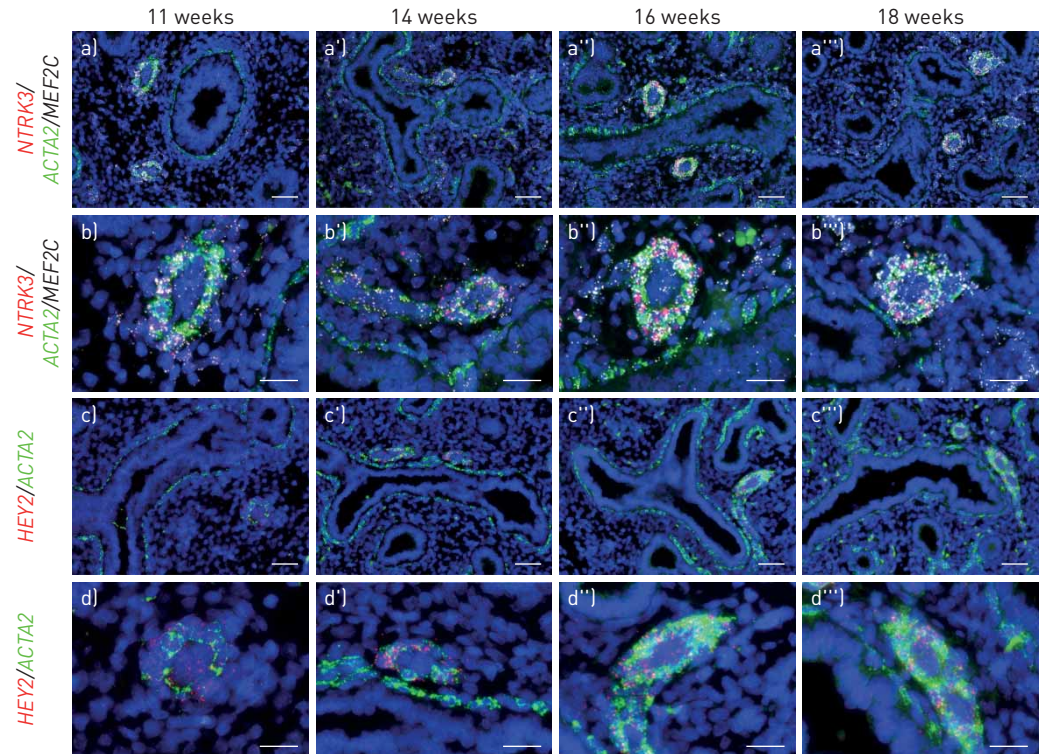
FIGURE 8 Vascular smooth muscle cell (VSMC) spatial validation. FISH: fluorescence *in situ* hybridisation; ASMC: airway smooth muscle cell. a–a''') FISH of neurotrophic receptor tyrosine kinase 3 (*NTRK3*; red) and myocyte enhancer factor 2C (*MEF2C*; white) in conjunction with α_2-smooth muscle actin (*ACTA2*; green) on human fetal lung sections at different gestational stages demonstrates that NTRK3 is localised in the VSMCs with no expression in the ASMCs, whereas MEF2C is predominantly expressed in the VSMCs with minimal expression in the ASMCs throughout development. b–b''') High magnification panels to confirm co-localisation of NTRK3 and MEF2C within the VSMCs. c–c''') FISH of Hairy/enhancer-of-split related with YRPW motif protein 2 (*HEY2*; red) and *ACTA2* (green) displayed that HEY2 is solely expressed in the VSMC with some expression in the proximal epithelium, but not expression in the ASMCs. d–d''') High magnification panels to confirm co-localisation of HEY2 within the VSMCs. Scale bar: a, c): 50 µm; b, d) (high magnifications): 25 µm. n=3 for each gestational stage.

Moreover, *MYLK* expression is increased in COPD samples, suggesting that this increase could be associated with ASMC hyperplasia [35]. Additional studies will be required to determine the timing at which these populations become distinct in the human fetal/embryonic lung, whether marker expression continues in a unique pattern through late lung development and adulthood, and if these same markers define these spatially separate populations in other mammals.

Conflict of interest: None declared.

References

1   Chung MI, Hogan BLM. Ager-CreER[T2]: a new genetic tool for studying lung alveolar development, homeostasis, and repair. *Am J Respir Cell Mol Biol* 2018; 59: 706–712.

2   Yang Y, Riccio P, Schotsaert M, *et al.* Spatial-temporal lineage restrictions of embryonic p63[+] progenitors establish distinct stem cell pools in adult airways. *Dev Cell* 2018; 44: 752–761.

3   Mori M, Mahoney JE, Stupnikov MR, *et al.* Notch3-Jagged signaling controls the pool of undifferentiated airway progenitors. *Development* 2015; 142: 258–267.

4   Jain R, Barkauskas CE, Takeda N, *et al.* Plasticity of Hopx[+] type I alveolar cells to regenerate type II cells in the lung. *Nat Commun* 2015; 6: 6727.

5   Lange AW, Sridharan A, Xu Y, *et al.* Hippo/Yap signaling controls epithelial progenitor cell proliferation and differentiation in the embryonic and adult lung. *J Mol Cell Biol* 2015; 7: 35–47.

6   Volckaert T, Yuan T, Chao CM, *et al.* Fgf10-Hippo epithelial-mesenchymal crosstalk maintains and recruits lung basal stem cells. *Dev Cell* 2017; 43: 48–59.

7   Kumar ME, Bogard PE, Espinoza FH, *et al.* Mesenchymal cells. Defining a mesenchymal progenitor niche at single-cell resolution. *Science* 2014; 346: 1258810.

8   Endale M, Ahlfeld S, Bao E, *et al.* Temporal, spatial, and phenotypical changes of PDGFRalpha expressing fibroblasts during late lung development. *Dev Biol* 2017; 425: 161–175.

9   Li R, Bernau K, Sandbo N, *et al. Pdgfra* marks a cellular lineage with distinct contributions to myofibroblasts in lung maturation and injury response. *Elife* 2018; 7: e36865.

10   Zepp JA, Zacharias WJ, Frank DB, *et al.* Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* 2017; 170: 1134–1148.

11   Danopoulos S, Alonso I, Thornton ME, *et al.* Human lung branching morphogenesis is orchestrated by the spatiotemporal distribution of ACTA2, SOX2, and SOX9. *Am J Physiol Lung Cell Mol Physiol* 2018; 314: L144–L149.

12   Miller AJ, Hill DR, Nagy MS, *et al. In vitro* induction and *in vivo* engraftment of lung bud tip progenitor cells derived from human pluripotent stem cells. *Stem Cell Reports* 2018; 10: 101–119.

13   Nikolic MZ, Caritg O, Jeng Q, *et al.* Human embryonic lung epithelial tips are multipotent progenitors that can be expanded *in vitro* as long-term self-renewing organoids. *Elife* 2017; 6: e26575.

14   Danopoulos S, Thornton ME, Grubbs BH, *et al.* Discordant roles for FGF ligands in lung branching morphogenesis between human and mouse. *J Pathol* 2019; 247: 254–265.

15   Sakao S, Taraseviciene-Stewart L, Wood K, *et al.* Apoptosis of pulmonary microvascular endothelial cells stimulates vascular smooth muscle cell growth. *Am J Physiol Lung Cell Mol Physiol* 2006; 291: L362–L368.

16   Yu W, Chen H, Yang H, *et al.* Dissecting molecular mechanisms underlying pulmonary vascular smooth muscle cell dedifferentiation in pulmonary hypertension: role of mutated caveolin-1 (Cav1[F92A])-bone marrow mesenchymal stem cells. *Heart Lung Circ* 2019; 28: 1587–1597.

17   Johnson PR, Roth M, Tamm M, *et al.* Airway smooth muscle cell proliferation is increased in asthma. *Am J Respir Crit Care Med* 2001; 164: 474–477.

18   Bandyopadhyay G, Huyck HL, Misra RS, *et al.* Dissociation, cellular isolation, and initial molecular characterization of neonatal and pediatric human lung tissues. *Am J Physiol Lung Cell Mol Physiol* 2018; 315: L576–L583.

19   Danopoulos S, Krainock M, Toubat O, *et al.* Rac1 modulates mammalian lung branching morphogenesis in part through canonical Wnt signaling. *Am J Physiol Lung Cell Mol Physiol* 2016; 311: L1036–L1049.

20   Xu Y, Mizuno T, Sridharan A, *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016; 1: e90558.

21   McCauley KB, Alysandratos KD, Jacob A, *et al.* Single-cell transcriptomic profiling of pluripotent stem cell-derived SCGB3A2+ airway epithelium. *Stem Cell Reports* 2018; 10: 1579–1595.

22   Xie T, Wang Y, Deng N, *et al.* Single-cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Rep* 2018; 22: 3625–3640.

23   Montoro DT, Haber AL, Biton M, *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 2018; 560: 319–324.

24   Wang Y, Tang Z, Huang H, *et al.* Pulmonary alveolar type I cell population consists of two distinct subtypes that differ in cell fate. *Proc Natl Acad Sci USA* 2018; 115: 2407–2412.

25   Du Y, Kitzmiller JA, Sridharan A, *et al.* Lung gene expression analysis (LGEA): an integrative web portal for comprehensive gene expression data analysis in lung development. *Thorax* 2017; 72: 481–484.

26   Stenmark KR, Durmowicz AG, Roby JD, *et al.* Persistence of the fetal pattern of tropoelastin gene expression in severe neonatal bovine pulmonary hypertension. *J Clin Invest* 1994; 93: 1234–1242.

27   Mariani TJ, Reed JJ, Shapiro SD. Expression profiling of the developing mouse lung: insights into the establishment of the extracellular matrix. *Am J Respir Cell Mol Biol* 2002; 26: 541–548.

28   Tan SY, Krasnow MA. Developmental origin of lung macrophage diversity. *Development* 2016; 143: 1318–1327.

29   Lynch TJ, Anderson PJ, Rotti PG, *et al.* Submucosal gland myoepithelial cells are reserve stem cells that can regenerate mouse tracheal epithelium. *Cell Stem Cell* 2018; 22: 653–667.

30   Werner P, Paluru P, Simpson AM, *et al.* Mutations in *NTRK3* suggest a novel signaling pathway in human congenital heart disease. *Hum Mutat* 2014; 35: 1459–1468.

31   Donovan J, Kordylewska A, Jan YN, *et al.* Tetralogy of Fallot and other congenital heart defects in *Hey2* mutant mice. *Curr Biol* 2002; 12: 1605–1610.

32   Fischer A, Klamt B, Schumacher N, *et al.* Phenotypic variability in *Hey2*[−/−] mice and absence of *HEY2* mutations in patients with congenital heart defects or Alagille syndrome. *Mamm Genome* 2004; 15: 711–716.

33   Qiao XH, Wang F, Zhang XL, *et al. MEF2C* loss-of-function mutation contributes to congenital heart defects. *Int J Med Sci* 2017; 14: 1143–1153.

34   Hobbs BD, de Jong K, Lamontagne M, *et al.* Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017; 49: 426–432.

35   Rouillard AD, Gundersen GW, Fernandez NF, *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016; 2016: baw100.