

## **Title**

**Single-cell RNA sequencing reveals novel cell differentiation dynamics during human airway epithelium regeneration**

## **Authors:**

Sandra Ruiz Garcia<sup>1†</sup>, Marie Deprez<sup>1†</sup>, Kevin Lebrigand<sup>1</sup>, Agnès Paquet<sup>1</sup>, Amélie Cavard<sup>1</sup>, Marie-Jeanne Arguel<sup>1</sup>, Virginie Magnone<sup>1</sup>, Ignacio Caballero<sup>2</sup>, Sylvie Leroy<sup>1,3</sup>, Charles-Hugo Marquette<sup>3</sup>, Brice Marcet<sup>1</sup>, Pascal Barbry<sup>1</sup>, Laure-Emmanuelle Zaragosi<sup>1</sup>

## **Affiliations:**

<sup>1</sup>Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, France.

<sup>2</sup>ISP, INRA, Université Tours, 37380, Nouzilly, France.

<sup>3</sup>Université Côte d'Azur, CHU de Nice, Pulmonology Department, Nice, France.

<sup>†</sup>These authors contributed equally

Correspondence to: [barbry@ipmc.cnrs.fr](mailto:barbry@ipmc.cnrs.fr)

## **Abstract**

### **Background:**

It is usually considered that the upper airway epithelium is composed of multiciliated, goblet, secretory and basal cells, which collectively constitute an efficient first line of defense against inhalation of noxious substances. Upon injury, regeneration of this epithelium through proliferation and differentiation can restore a proper mucociliary function. However, in chronic airway diseases, the injured epithelium frequently displays defective repair leading to tissue remodeling, characterized by a loss of multiciliated cells and mucus hyper-secretion. Delineating drivers of differentiation dynamics and cell fate in the human airway epithelium is important to preserve homeostasis.

### **Results:**

We have used single cell transcriptomics to characterize the sequence of cellular and molecular processes taking place during human airway epithelium regeneration. We have characterized airway subpopulations with high resolution and lineage inference algorithms have unraveled cell trajectories from basal to luminal cells, providing markers for specific cell populations, such as deuterosomal cells, i.e. precursors of multiciliated cells. We report that goblet cells, like secretory cells, can act as precursors of multiciliated cells. Our study provides a repertoire of molecules involved in key steps of the regeneration process, either keratins or components of the Notch, Wnt or BMP/TGF $\beta$  signaling pathways. Our findings were confirmed in independent experiments performed on fresh human and pig airway samples, and on mouse tracheal epithelial cells.

### **Conclusions:**

Our single-cell RNA-seq study provides novel insights about airway epithelium differentiation dynamics, clarifies cell trajectories between secretory, goblet and multiciliated cells, identifies novel cell subpopulations, and maps the activation and repression of key signaling pathways.

## **Keywords**

Airway epithelium, regeneration, single-cell RNA-seq, differentiation, multiciliated cells, secretory cells, goblet cells, mucus-secreting cells, basal cells, keratins, pathways, clustering.

## **Background**

In mammalian airways, a pseudo-stratified mucociliary epithelium constitutes an efficient first line of defense of the respiratory tract against a large panel of inhaled substances. This epithelium forms a complex ecosystem, mainly composed by: multiciliated cells (MCCs), projecting hundreds of motile cilia at their apical surface, goblet cells (GCs), secreting protective mucins on the luminal surface, secretory cells (SCs), producing anti-microbial and anti-inflammatory peptides, and basal cells, playing a role in adhesion and stability of the epithelium [1,2]. Altered balance between multiciliated and goblet lineages (i.e. decreased number of MCCs with increased number of GCs) is a hallmark of chronic respiratory diseases, such as chronic obstructive pulmonary disease, primary ciliary dyskinesia, asthma or cystic fibrosis. These diseases, which collectively affect hundreds of millions of people, are characterized by frequent injuries, repair defects, tissue remodeling and altered mucociliary clearance [3–5]. In order to accurately characterize the cellular and molecular identity of progenitor cells of the airway epithelium that contribute to efficient tissue regeneration and of the mechanisms that control the MCC/GC balance, key facts about lung epithelial cell fates have been investigated in mouse, mainly injuring the epithelium with noxious agents. Lineage tracing studies have identified basal cells as the main airway progenitor cells, as they display self-renewal capacities and ability to differentiate into MCCs and SCs [2,6,7]. Basal cells in mouse are abundant in the trachea and in the main bronchi, but they are absent in smaller airways [8]. In human, they populate the whole airways, though showing a 2-fold decrease in numbers in smaller airways, i.e. with diameters below 0.5mm [9]. The TP63 transcription factor, keratins 5 and 14 (KRT5 and KRT14), podoplanin (PDPN), nerve growth factor receptor (NGFR), galectin 1 (LGALS1), integrins  $\alpha 6$  and  $\beta 4$  (ITGA6 and ITGB4), laminins  $\alpha 3$  and  $\beta 3$  (LAMA3 and LAMB3) have been frequently used as selective markers of BCs [2,6]. While a direct differentiation of a subpopulation of BCs into MCCs has been reported upon injury [10], the consensus emerging from lineage tracing studies suggests that BCs usually differentiate first into SCs [11]. SCs, known as club (or Clara) cells, are widespread from mouse trachea to bronchioles, and are also found in the human airways, though at a much less abundance, being absent from the upper airways and enriched in the terminal and respiratory bronchioles [12]. SCs are detected at a luminal position of the epithelium, showing a characteristic columnar shape, being filled with secretory granules containing anti-microbial, anti-inflammatory peptides [13] and contributing to xenobiotic metabolism [14]. The most extensively studied marker of SCs is the anti-inflammatory secretoglobulin SCGB1A1. Lineage tracing studies have used SCGB1A1 to determine the fate of SCs and have shown that they give rise to MCCs, identified by the specific expression of the transcription factor FOXJ1 [11,15] and to GCs, identified by the expression of specific mucin MUC5AC [2,16]. Molecular mechanisms regulating cell fate decisions in the airway epithelium lineages have been investigated over the past few years and

they have established the role of Notch signaling during commitment of BCs. It is now clear that Notch activation leads to the SC/GC lineages as Notch inhibition leads to commitment towards the MCC fate [17–20]. In agreement with these findings, our group has shown that Notch pathway inhibition by the miR-34/449 families of microRNAs is required for MCC differentiation [21,22]. Although *in vivo* lineage tracing studies have largely unveiled cell lineage hierarchies in the airway epithelium, they have some inherent limitations. They can only be performed in mouse, after specific, drastic, non-physiological forms of injuries which may not completely reveal physiological tissue turnover. Moreover, they rely on genetic cell labelling, usually *Krt5* for BCs and *Scgb1a1* for SCs, thus, they are not necessarily comprehensive and cannot provide a full picture of cell hierarchies in the airway epithelium. In human, there is no possibility to perform such studies, and cell lineage hierarchies must be inferred indirectly by *in vitro* approaches. Single-cell RNA-sequencing, coupled with computational methods measuring transition between cell states, has emerged as a complementary and powerful approach to reveal lineage hierarchies in many tissues [23–25] and even in a whole organism, by capturing undifferentiated, intermediate and terminally differentiated states [26]. In the lung, one pioneer single-cell study on 198 cells has delineated lineage hierarchies of alveolar cells in the mouse distal lung [27]. Recent studies have established first atlases of the airways in mouse [28] and human [29,30]. However, extensive characterization of cell population diversity and cell lineages in the human airways remains to be achieved. Here, we have performed and collected single-cell RNA-seq data from fresh human airway epithelial tissues and all along an experiment of 3D *in vitro* regeneration of human airway epithelium. Our work has resulted in a comprehensive cell trajectory roadmap of human airways, which identifies novel cell populations and offer new insights into molecular mechanisms taking place during the mucociliary epithelium regeneration.

## **Results**

### ***Reconstruction of cell lineage in regenerating airway epithelium by single-cell RNA-seq***

To identify trajectories of cells composing the human airway epithelium, we have analyzed single-cell transcriptomes obtained at successive stages during *in vitro* 3D regeneration of this tissue (Fig. 1a, b). We first validated that this *in vitro* model faithfully recapitulated native airway tissues by comparing cell population compositions of 3D regenerated airway tissues (Human Airway Epithelial Cells, HAECs) with that of fresh human airway tissues. We performed single-cell RNA-seq of epithelial cells dissociated from nasal brushing samples and from nasal turbinates which were resected during surgery, as well as single-cell RNA-seq of HAECs at late time point of *in vitro* air-liquid interface differentiation (3D cells). Our main results were obtained with HAECs that were cultured in a Pneumacult media (StemCell Technologies), which allows the production of multiciliated and goblet

cells, but additional experiments were also performed with HAECs cultured in BEGM (Lonza), which rather favors the production of multiciliated cells. Cell identity was inferred from the expression of specific marker genes, such as *KRT5* and *TP63* for basal cells (BCs), *SCGB1A1* for secretory cells (SCs), *MUC5AC* for goblet cells (GCs) and *FOXJ1* for multiciliated cells (MCCs). These cell types were robustly found in all samples at various proportions, validating the use of these *in vitro* models to trace airway cell lineages (Supplemental Figure S1a-c). We also confirmed that cell type proportions inferred from single-cell RNA-seq were correlated with cell type proportions inferred from protein measurements by performing immunostaining of selected population markers (Supplemental Figure S1d, e). In addition, we have evaluated the effect of cell dissociation on gene expression (Supplemental Figure S2a). We found that cell dissociation did not produce a major impact on gene expression with the exception of *FOS* and *FOSB* which were highly upregulated (Supplemental Figure S2b, c). Molecular function enrichment with Ingenuity Pathway Analysis (Qiagen) showed that “cell death and survival” and “cellular growth and proliferation” were the only molecular functions to be regulated with a p-value inferior to 0.001 (Supplemental Figure S2d).

We have generated single-cell transcriptomes of HAECs cultivated in Pneumacult medium at 3 time points of *in vitro* airway regeneration (ALI7, ALI12, and ALI28) (Fig. 1b). Time points were chosen as most representative of the known steps of airway regeneration: proliferation, polarization and specification [31]. This experiment was complemented by 6 additional time points of HAECs cultivated in BEGM medium (ALI2, ALI4, ALI7, ALI12, ALI17 and ALI22). In a first approach, each time point was analyzed independently. We did 10 random selections of cells, corresponding to subgroups containing 90% of the initial number of cells. The resulting gene expression submatrices were then iteratively clustered (10 times with varying parameters), and a census was applied to define the most robust cell types. We then studied the variations of these populations during the entire time course. Cells clustered in 6 main populations in the Pneumacult condition: (1) cycling (*MKI67*+) and (2) non-cycling (*MKI67*-) BCs (*KRT5*+/ *TP63*+), (3) supraBCs (*KRT5*+/ *TP63*- / *KRT13*+/ *KRT4*+), (4) SCs (*SCGB1A1*+), (5) GCs (*MUC5AC*+) and (6) MCCs (*FOXJ1*+) (Figure 1c). Cell population proportions evolved during the time course, with a global reduction of BCs and SCs, a first detection of supraBCs at ALI7, followed by an increase of the proportion of this cell population at ALI28 and a first detection of GCs and MCCs at ALI28 (Fig. 1d). In the BEGM condition, cells clustered in 7 cell populations (Supplemental Figure S3a, b). We did not detect SCs and GCs with this culture condition, but we found instead a cell population that we termed “Secretory-like cells”, given their high gene expression similarity with SCs, except for *SCGB1A1*, which was not detected (Supplemental Figure S4). Additional cell types were found in these samples: *KRT5*- supraBCs (*TP63*-/ *KRT13*+/ *KRT4*+) and 2 cell populations that we termed as “undefined intermediates 1” and “undefined intermediates 2” because their gene expression profile did not allow

unambiguous classification. Differential gene expression showed that these undefined intermediates 1 and 2 expressed specific genes, such as *SPINK1* and *SPINK7*, respectively. A representation of all cells analyzed in each culture was generated using t-SNE graphs on the aggregate of all time points for each medium condition (Fig. 1e for Pneumacult and Supplemental Figure S3c for BEGM). Cell trajectories and transitions from one cell population to another were deduced from a trajectory inference analysis using Monocle 2, followed by differential expression analysis between consecutive cell states in pseudotime using Seurat. In the BEGM condition, a unique cell trajectory was found (Supplemental Figure S3d), starting with cycling and non-cycling BCs at its beginning, followed by *KRT5*<sup>+</sup> and then *KRT5*<sup>-</sup> supraBCs cells, and with MCCs at its end. Before reaching the secretory-like cell state, the trajectory goes through the 2 “undefined intermediate” cell states. Despite the absence of *SCGB1A1* expression in secretory-like cells (*SCGB1A1*<sup>-</sup>/*BPIFA1*<sup>+</sup>/*KRT8*<sup>+</sup>), these cells were ordered in the pseudotime before MCCs, as expected for canonical SCs (Supplemental Figure S3d-f). A more complex trajectory was observed with the Pneumacult condition, in which Monocle2 detected a bifurcation into 2 distinct branches after the SC stage: the larger branch leading to *FOXJ1*<sup>+</sup> MCCs, and the smaller one leading to *MUC5AC*<sup>+</sup> GCs (Fig. 1f, g). A closer examination of the pseudotime ordering and differential gene expression (Fig. 1h) revealed that few *MUC5AC*<sup>+</sup> cells were found on the MCC branch, after the GC bifurcation and that some *FOXJ1*<sup>+</sup> cells retained expression of *MUC5AC*. Altogether, our findings confirm SCs as precursors of both MCCs and GCs. They also suggest that GCs can also act as MCC precursors in airway epithelial regeneration.

### ***Goblet cells can act as differentiation intermediates for multiciliated cells***

We further investigated the possibility that some GCs may correspond in fact to precursor cells for MCCs. A first evidence is coming from our robust clustering analyses, either when they were performed on cells coming from *in vitro* samples or from fresh tissues. The two populations of GCs and SCs displayed very similar gene expression profiles, and were discriminated based on their levels of expression of *MUC5AC* and *MUC5B* which were higher in GCs (Supplemental Figure S1a-c). In the Pneumacult experiment, 24 out of the 54 top genes for GCs were also associated with SCs (Fig. 2a). Among these transcripts was *SCGB1A1*, while expression of *MUC5AC* and *MUC5B* was more robust in GCs (Fig. 2b). When we directly assessed differential gene expression between cells located at the two ends of the GC branch (Fig. 1f, Fig. 2c), we confirmed the high similarity of gene expression existing between SCs and GCs. GCs differed from SCs by higher levels of expression for SC/GC specific genes. This was observed for molecules such as mucins (*MUC1*, *MUC4*, *MUC5B*, *MUC5AC*), secretoglobins (*SCGB1B1* and *SCGB3A1*), PLUNC antimicrobial factors (*BPFA1* and *BPIFB1*) and *SLPI*, the secretory

leukocyte protease inhibitor (Fig. 2c). These properties led us to consider GCs as ‘hyperactive’ SCs and led to the prediction that these cells could also function as MCC precursors. This point was tested by quantifying the cellular expression of *MUC5AC* and *FOXJ1*, and by measuring the percentage of doubly labeled cells. Our rationale was that a detection of cells that express at the same time *MUC5AC* and *FOXJ1* would suggest the existence of a transitory state between GCs and MCCs. Fig. 2d, 2g and 2j indeed show that 8.9% of GCs + MCCs express at the same time *MUC5AC* and *FOXJ1*. It also shows the existence of SCs/MCCs expressing both *SCGB1A1* and *FOXJ1*, which correspond to a more conventional type of precursors for MCCs (Fig. 2m). The presence of *MUC5AC*+/*FOXJ1*+ and *SCGB1A1*+/*FOXJ1*+ cells was not restricted to a cell culture model, as these transitional cells were also detected in fresh biopsies from human bronchi (Fig. 2e, h, k, n) and pig trachea (Fig. 2f, i, l, o). The presence of doubly-labeled cells was confirmed by qPCR in a fully independent dataset, derived from a HAEC culture (Supplemental Figure S5). In that other experiment, we isolated the cells with the C1 technology (Fluidigm) and quantified gene expression by quantitative RT-PCR with a Biomark (Fluidigm). Cells isolated with the C1 were visually inspected, and these experimental settings ensured the absence of cell doublets. We found that 4 cells out of 74 expressed GC specific genes (namely *MUC5AC*, *MUC5B* and *TFF3*), together with MCC specific genes (*FOXJ1*), and more specifically, immature MCC genes (*PLK4*, *MYB* and *CDC20B*) [32] (Supplemental Figure S5a, b). The result was further confirmed after reanalyzing a recent dataset published by others [29] (Supplemental Figure S5c). A further confirmation came from the detection at a protein level of cells that were labeled at the same time by *MUC5AC* and acetylated Tubulin, a specific protein marker of the cilia (Fig. 2p). A final point came after a survey of our data with “RNA velocity” [33]. Velocity can predict the fate of individual cells at a timescale of hours by distinguishing the expression of spliced and unspliced forms of transcripts. We used Velocity to analyze the behavior of three transcripts: *CEP41*, a specific marker of cells in an early phase of multiciliated differentiation, *SCGB1A1* and *MUC5B*. RNA velocity algorithm calculates a residual value of each gene, which indicates expected upregulation when it is positive and expected downregulation when it is negative. Positive residuals were found for transcripts of *CEP41* in the GC population, predicting an upregulation of this gene’s expression in the next hours. A different picture was observed for the transcripts of *SCGB1A1* and *MUC5B*, in which negative residuals were found in the GC and SC populations, indicating an expected downregulation of the corresponding transcripts in the next hours (Fig. 2q). Altogether, these data indicate that GCs can act as precursors for MCCs in normal *in vitro* and in homeostatic *in vivo* airway regeneration.

***Refining cell clustering identifies 6 additional clusters, including a discrete population of pre-MCC “deuterosomal” cells***



To gain further insight in the diversity of cell populations composing the airway epithelium and the transitional ones occurring during the regeneration, we considered additional clusters that could be derived from our sub-clustering analysis, by accepting less discriminations between them than between the initial 6 previously identified clusters. This deeper analysis defined 12 clusters, rather than the 6 initial ones (Fig. 3a, Supplemental Figure S6a). The non-cycling BC population was split into 2 clusters that we termed BC1 and BC2. The major difference between these 2 clusters was the higher level of expression for genes associated with cell migration: *FN1*, *VIM*, *SPARC* and *TAGLN* in the BC2 cluster. Analysis of enriched canonical pathways with Ingenuity Pathway Analysis showed enrichment for integrin, actin cytoskeleton and Rho GTPase signaling as well as the pathway “regulation of actin-based motility” in BC2 compared to BC1, suggesting an increased migratory activity in BC2 (Supplemental Figure S7). Since BC2 cells were associated with higher pseudotime values than BC1, BC2 likely correspond to the population of most differentiated BCs, just before reaching a suprabasal state. The *supraBC* and SC populations could also be further split into 3 distinct sub-populations, resulting in a total of 3 new populations of *supraBC* and 3 new populations of SC (Fig. 3a and Supplemental Figure S6a). Each of them displayed its own distinct gene set enrichment (Supplemental Figure S7). Among them, the SC2 subpopulation appeared particularly interesting, since it displayed a strong enrichment score for the feature “immune cell migration, invasion and chemotaxis”, and also a strong positive enrichment for canonical pathways such as “neuroinflammation signaling” and “dendritic cell maturation”. This was explained by the increased gene expression of targets for pro-inflammatory molecules such as TNF, IFNG, NFkB, IL1A/B, IL2, or IL6, as well as decreased gene expression for targets for the anti-inflammatory PPARG pathway (Supplemental Figure S7). This may confer to this subpopulation of SCs a unique relationship with the immune response. The MCC cluster, containing *FOXJ1*+ cells, was further split in 2 discrete clusters: (1) the largest one is positive for mature MCC genes such as *DNAH5*, and corresponds to terminally differentiated MCCs; (2) the second one specifically expresses several molecules that are important for the biosynthesis of hundreds of motile cilia. Among them is *DEUP1*, a hallmark of massive centriole amplification at deuterosomes (Fig. 3b). This led us to coin these cells “deuterosomal” cells. This subpopulation is clearly distinct from mature MCCs in our clustering analysis (Fig. 3b) and expresses also highly specific markers such as *PLK4*, *CCNO*, and *CEP78* (Supplemental Figure S8). We have confirmed the existence of deuterosomal cells in three distinct species with single cell experiments performed on mouse tracheal epithelial cells (MTECs) dissociated at ALI3 (i.e. the time point of maximal centriole amplification in this model), newborn pig trachea and human bronchial biopsy from a healthy adult subject (Fig. 3c). In all samples, even under homeostatic conditions, we noticed the presence of deuterosomal cells that clustered independently of mature MCCs. This deuterosomal cell population was characterized by the expression of gene markers, either uniquely expressed in deuterosomal cells, or also expressed in MCCs and cycling BCs



(Fig. 3d). Our analysis shows that the group of deuterosomal cells specifically expressed 149 genes, while sharing 33 and 244 genes with cycling BCs and mature MCCs, respectively (Fig. 3e). Among the 33 genes in common with cycling BCs, we noticed the re-expression of several cell cycle-related genes, which are required for the massive amplification of centrioles that takes place [34,35]. The most specific genes are displayed in Fig. 3d. This analysis not only confirmed the known expression of *CDK1* in deuterosomal cells [35], as it also highlights the expression in deuterosomal cells of genes coding for centromere proteins (*CENPF*, *CENPU*, *CENPW*), securin (*PTTG1*), a core subunit of the condensin complex (*SMC4*) and cyclin-dependent kinases regulatory subunits (*CKS1B*, *CKS2*). We confirmed the deuterosomal-specific expression of *CDC20B*, the miR-449 host gene. We have recently shown that *CDC20B* is a key regulator of centriole amplification by deuterosomes [32]. Incidentally, we noticed the existence of a novel and short isoform of this gene, arising from alternative splicing and results in the inclusion of a novel exon after the location of the miR-449 family (Fig. 3b and Supplemental Figure S9a, b). We found that this short *CDC20B* isoform was also detectable in mouse RNA-seq data (Supplemental Figure S9c). Comparison of transcript abundance in several samples, including the Pneumacult ALI28 and the human bronchial biopsy showed higher levels for short *CDC20B* compared to *CDC20B* (Supplemental Figure S9d). This short *CDC20B* likely corresponds to the major source of miR-449 in deuterosomal cells. A list of novel markers of deuterosomal cells that are specifically expressed in this cell population is provided in Supplemental Table S1. Some of these genes had never been described before in the context of centriole amplification, such as the Yippee-like factor *YPEL1* or the Notch pathway related hairy-enhancer-of-split family of transcription factors *HES6* (Supplemental Figure S8). Gene set enrichment of the deuterosomal population specific genes (Fig. 3f) shows an enrichment for terms such as cilium assembly, centrosome maturation but also, cell-cycle mechanism-related terms such as resolution of sister chromatid cohesion, regulation of *AURKA* and *PLK1* activity, *CDH1* autodegradation. “Mitochondrial membrane part” was also among the enriched terms, suggesting an increase in mitochondria numbers at this stage. This deuterosome-specific signature perfectly delineates the events that are occurring at this MCC differentiation stage and provides an extensive repertoire of specific cell-cycle related genes which are re-expressed at the deuterosomal stage. This deuterosomal population is a consistent population, much larger than recently described rare cell populations such as ionocytes [28,29], which we also identified (Supplemental Figure S6c).

### ***Establishing a keratin switch pattern during airway regeneration***

An extremely diverse repertoire of keratins has been associated with different types of epithelial cells. Considering that expression of specific keratins varies with cell type, period of embryonic

development, stage of histologic differentiation, cellular growth environment, and disease state, we reasoned that it would be useful to compare our scRNA-seq data with the expression of different keratins. *KRT5* and *KRT14*, are largely used as BC markers in the airway and lung epithelia, and also in other epithelia such as bladder [36], prostate [37] or mammary gland [38]. *KRT8* is clearly associated with luminal cell types [6]. Besides, there is no clear repertoire of the other keratins that are associated with airway cell types. A recent study performed on mouse and human models of *in vitro* regeneration identified *KRT4* and *KRT13* in a subpopulation reminiscent of our supraBCs, as it emerges between BCs and SCs [29]. We have established an extensive single-cell repertoire of KRT expression during airway regeneration based on pseudotime ordering in our Pneumacult ALI28 and pig trachea datasets. Our analysis confirmed in both datasets the presence of *KRT5* and *KRT14* in BCs, of *KRT4* and *KRT13* in supraBCs, and the expression of *KRT8* in luminal cell types (SCs, GCs and MCCs) (Fig. 4a and e). We consistently noticed that expression profiles of *KRT13* and *KRT4* did not completely overlap, with *KRT13* detected at earlier pseudotimes than *KRT4*. This was confirmed at a protein level by a quantification by immunostainings of the proportion of *KRT5*+/*KRT13*+ and *KRT5*+/*KRT4*+ double positive cells (Fig. 4b). Fig. 4c shows that there was more *KRT5*+/*KRT13*+ (7.4%) than *KRT5*+/*KRT4*+ (4.9%) double positive cells, which is consistent with an earlier expression of *KRT13* compared to *KRT4*. In the pig trachea, we also found a very clear shift, with 16.8% and 11.2% of *KRT5*+/*KRT13*+ and *KRT5*+/*KRT4*+ double positive cells, respectively (Fig. 4d and Supplemental Figure S10). Our results show that *KRT4* and *KRT13* are not strictly expressed at the same time during airway regeneration and their expression delineate subtle differences in cell subpopulations. Additional keratins such as *KRT16* and *KRT23* displayed a specific supraBC expression (Fig. 4e). We also identified additional keratins that were more specifically associated to more differentiated cell types: *KRT7* and *KRT18* were strongly enriched in SCs, but their expression completely dropped in MCCs, while *KRT8* was still expressed (Fig. 4e). Expression patterns for these cell type specific keratins were confirmed in the pig trachea (Supplemental Figure S10c). The keratin switch pattern is indeed sufficiently specific to allow a reconstruction of the cell trajectories during airway regeneration.

### ***Establishing a combinatorial repertoire of signaling pathways during airway regeneration***

We have finally analyzed the cell specificity of expression of important signaling pathways, to catch mutual influences between distinct cells that could play a role in airway regeneration. Our investigation was focused on the Notch, BMP/TGF $\beta$  and Wnt pathways. For each different component, we classified them as ligands, receptors, or targets. The expression profiles are shown as heatmaps, cells being sorted by their subgroups, then by ascending pseudotime value (Fig. 5a-c).

Notch pathway (Fig. 5a). BCs express the ligands *DLL1*, *JAG1* and *JAG2*, as well as the receptor *NOTCH1*, as expected [6,29]. In this population, no target gene expression is detected, suggesting that the pathway is inactive. Interestingly, BCs also express *LFNG*, which has been described to inhibit JAG1 signaling via NOTCH1 [39]. SupraBCs cells express *NOTCH1*, *JAG1* and *JAG2* but, as opposed to BCs, they show clear activation of the Notch pathway by expression of the target genes *HEY1*, *HES2* and *HES4*. *NOTCH3* expression is turned on and is specific to this population. In SCs/GCs, *NOTCH2* is the major receptor to be detected and signal activation remains as evidenced by the expression of *HEY1* and *HES4*. SCs/GCs also express the non-canonical Notch ligand *NTN1*. In Deuterosomal cells/MCCs, a clear shift is observed. *NOTCH2*, *NOTCH3*, *HEY1* and *HES4* are turned down, and *NOTCH4* is specifically expressed. As previously described, *JAG2* [29], which is present in BCs, then absent in supraBCs and SCs/GCs, is re-expressed in the MCC compartment. We have found the same behavior for *DLL1* and the non-canonical ligand *DNER1*. Thus, MCC express some Notch ligands. Strikingly, a major inhibitory signature dominates in MCCs, with the expression of *CIR1* and *SAP30*, which are transcriptional co-repressors, as well as of *DYRK1A*, an inhibitor of NICD. *HES6*, which expression is not regulated by Notch signaling but which has been identified as a Notch pathway inhibitor through its binding to HES1 [40], displays an expression pattern that is highly enriched in deuterosomal cells (Fig. 5a and Fig. 3d).

Wnt pathway (Fig. 5b). The Wnt target genes *SNAI2* and *TCF4*, indicators of an active pathway are mainly enriched in the BC population, especially in BC2 for *SNAI2*. In the BC population, *WNT10A* and *LRP1* are strongly enriched, and several *SOX* family members (*SOX2* and *SOX21*) are underrepresented, especially in the cycling BCs, suggesting an activation of the pathway in this compartment. In the MCC population, the situation is more complex. Despite a slight expression of *TCF4* together with positive regulators of the pathway such as *WNT9A*, *FZD6*, *APPL2*, *CSNK1G1* (a casein kinase component that can act as an activator or inhibitor of the pathway [41]), no *SNAI2* expression is detected and known repressors of the Wnt pathway are also overrepresented. Indeed, MCCs express significant levels of the transcriptional repressors *SOX2*, *SOX21* and display strong enrichments for the Reptin components *RUVBL1* and *RUVBL2* (Fig. 5b).

BMP/TGF $\beta$  pathway (Fig. 5c). BMP ligands, such as *BMP2*, and *BMP7*, are both enriched in the BC population, while *BMP3* and *BMP4* are both enriched in the SC/GC populations. A different picture was found for BMP receptors, for which we did not found specific cell population expression. The most striking observation is the specific expression in BCs of BMP inhibitors such as the inhibitory ligand *FST* and the intracellular inhibitor *FKBP1A* (also known as *FKBP12*). As BMP signaling is considered as a brake for proliferation, inhibition in the BC compartment is consistent with the maintenance of a proliferative potential of this progenitor population. Regarding the TGF $\beta$  pathway, a clear signal of activation is detected in the deuterosomal/MCC population, with specific expression of the target

genes *SERPINE1* (PAI-1), *CTGF*, *ATF3*, *TGFBR3* and *IRF7*, consistent with the previous finding that TGF $\beta$  pathway regulates motile cilia length by affecting the transition zone of the cilium [42]. We did not detect TGF $\beta$  ligands in the MCC population but rather found them expressed in BCs (TGFB1) and supraBCs (TGFB3).

We have confirmed the main distribution of the three pathway components sample differentiated with the BEGM medium (Supplemental Figure S11) and in 2 fresh tissue samples (human bronchial biopsy and nasal turbinate) for which a selection of genes is shown in Fig. 5d. Collectively, our data provide for the first time a detailed repertoire of signaling pathways at work during airway regeneration, with receptors and ligands specifically expressed at each cell stage.

## **Discussion**

In this study, we have established a comprehensive single-cell atlas throughout the entire time course of human airway regeneration *in vitro*. We carefully chose representative time points in the process and we quantified the proportion and identity of each cell population at different time points after the establishment of the air liquid interface. We provide here for the first time a comparison of the most widely used culture media in 3D culture of airway epithelial cells: BEGM, the most established commercial medium with which the majority of studies have been performed, and a more recent commercial medium, Pneumacult. In the BEGM medium, we have performed analyses at earlier time points, i.e. ALI2 and ALI4. These time points allowed us to measure the extent of cell proliferation during *in vitro* regeneration. Cycling basal cells account for approximately 40% of total cells at ALI2 and ALI4, and this number drops to 5% at ALI7. These early time points also showed that suprabasal cells appeared early in these conditions, as they were already detected at ALI4. With the BEGM medium, unlike with the Pneumacult medium, we never detected any goblet (*MUC5AC*+) nor “canonical” secretory (secretoglobin+) cell, even after long periods of time, and on several dozens of cultures coming from distinct donors (Supplemental Figure S1, S3, and personal observations). However, we found a cell population that we have termed “secretory-like”. “Secretory-like” cells express a gene pattern very similar to that of canonical secretory cells, and they can differentiate into MCCs. Interestingly, goblet cells were detected in BEGM medium after an IL-13 treatment [41, and our personal observations]. Future work should investigate whether secretory-like cells first evolve into canonical secretory cells and then goblet cells upon IL-13 treatment.

In the Pneumacult medium, but also in freshly dissociated human bronchial biopsy and pig trachea, we have detected cells expressing both *MUC5AC* and *FOXJ1*. We believe this finding is in line with our lineage inference and RNA velocity data showing that goblet cell can be precursors for MCCs. Other groups had previously detected cells expressing both these markers, in the context of GC

hyper/metaplasia induced by Sendai virus infection or after IL-13 treatment [44–46]. These findings led them claim that MCC can transdifferentiate into GCs. However, none of their data clearly showed a difference in the number of these double-positive cells between control and treated conditions. Our present finding probably provide a different answer to this conundrum, saying that these double-positive cells do exist in the absence of IL-13 stimulation. We noticed that their expression profiles posit them more naturally as precursors of MCCs than as trans-differentiated multiciliated cells. Turner and colleagues [45] proposed that MCCs could trans-differentiate into GCs after performing *in vitro* lentiviral transduction of HAECs with a vector containing a Cre recombinase under the control of the *FOXJ1* promoter. These findings were not confirmed by Rajagopal’s group who showed no GC arising from MCC in a context of OVA-induced mucous metaplasia in mouse airways, using *in vivo* lineage tracing with *Foxj1*-cre mice [47].

Our datasets are not limited to HAECs, but also provide useful information about mouse, human, and pig tissues. They correspond to *in vitro* MTECs, as well as cells freshly isolated from human nasal brushing, turbinate, bronchial biopsy and finally pig trachea. We have used this variety of samples to confirm our findings with HAECs, and to invalidate the possibility of culture artefacts. Our study was not focused on rare cell types such as pulmonary neuroendocrine, brush cells or ionocytes, which have been recently described elsewhere. We have indeed detected cells, which displayed the gene signature of ionocytes, characterized by expression of *CFTR* and other ion transporters, as well as specific transcription factors such as *ASCL3* or *FOXI1* [28,29]. Our investigation was rather focused on the diversity of the main cell types that compose the epithelium. We have identified 3 subtypes of basal cells among which cycling basal cells, and a group of basal cells expressing significantly higher levels of genes involved in extracellular matrix connection and actin-based motility. As these cells are the latest basal cells in our pseudotime, they likely represent an important transient migratory state on the way to the suprabasal state.

An additional peculiarity found within the secretory compartment comes from one of our three secretory subpopulations which displayed an immune-related gene signature. So far, diversity within the secretory compartment was established after expression of different members of the secretoblogin family [48] or via activation level of the Notch pathway [49]. We propose that diversity within the secretory cell compartment should also include specialized functions related to the interaction between the epithelium and immune cells. Additional experiments including protein labeling on fresh tissue sections have now to be performed in order to confirm this diversity and identify the spatial distribution of these subpopulations.

Our study has also provided a first extensive gene signature of the deuterosomal population, which play a key role during differentiation of multiciliated cells. In line with what has been shown recently by our group and others [32,35], cell-cycle related genes become re-expressed in this population on

non-cycling cells. We have confirmed the very specific expression of *CDC20B*, a key player of centriole amplification [32], and have identified, both in human and mouse, a novel isoform of this transcript which displays higher expression than the annotated long isoform. As the pre-mRNA corresponding to this short isoform comprises the *miR-449*-encoding intron, we suggest that this isoform should indeed be the major source of *miR-449* in deuterosomal cells. The alternative splicing that is responsible for this alternative isoform might represent an optimization of gene expression regulation to efficiently increase *miR-449* levels.

Besides the identification of each cell type, we have also characterized the distribution of some important signaling pathway components. First, we have focused on the Notch pathway as it is already known as a major regulator of the mucociliary differentiation. We have confirmed the distribution of ligands and receptors described by others [19,20,29,50]. We have also confirmed the absence of Notch activation in BCs and MCCs, with *HES4* being the most representative target gene in our model. BCs express *NOTCH1* and NOTCH ligands. However, no clear Notch pathway activation can be detected within this cell population even in an uneven manner as it could be expected from Notch lateral inhibition. This could be solely the result of weak *NOTCH1* expression but could also be a consequence of specific expression of Notch inhibitors such as the ligand *LFNG* or Casein kinase II subunit beta (*CSNK2B*) [51,52]. At the other end of mucociliary differentiation, inhibition of the Notch pathway has been widely documented in MCCs: failure to inactivate Notch results in failure of MCC signaling differentiation. Here we suggest a novel mechanism for Notch pathway inactivation by showing specific expression of several Notch transcriptional inhibitors at the deuterosomal stage. These include *HES6*, an inhibitory HES acting through HES1 binding [40,53], *DYRK1A*, an inhibitor of Notch Intracellular Domain transcriptional activity [54], as well as *CIR1* and *SAP30* which are transcriptional repressors of Notch/CSL transcriptional complex [55]. As opposed to MCCs and BCs, SCs must undergo clear Notch activation to maintain their cell identity, and differentiate into GCs [18–20]. However, the onset of this signal's activation has not been widely studied. Mori and colleagues have described *NOTCH3* expression in TP63-negative cells in a parabasal position of the epithelium, which likely correspond to the cells we, and others, have termed supraBCs [50]. We have confirmed that *NOTCH3* RNA is absent from BCs and becomes upregulated in supraBCs. We went further by showing *HES4* becomes expressed at this cell stage, confirming that Notch pathway activation starts at the supraBC stage. Thus, we emphasize here the importance of this intermediate cell population, though it has not been well characterized so far, for establishing Notch activation and subsequent differentiation.

Contrary to the Notch pathway, the Wnt/ $\beta$ -catenin pathway has not been extensively studied in the context of airway epithelium differentiation. Crosstalk between these 2 pathways has at least been suggested by other studies performed in other systems. For instance, in the hair follicle precortex,  $\beta$ -catenin stimulates Notch signaling by inducing *Jag1* transcription [56]. In the context of airway



epithelium, a very recent study demonstrated that  $\beta$ -catenin signaling is required for the early stages of mucus and ciliated cell differentiation, which they defined as “specification”, but was detrimental to the later “commitment” stages [57]. Wnt also seems to be related to epithelial remodeling upon inflammatory situations: Ordovas-Montanes and colleagues have recently shown that in nasal polyps, which are characterized by an inflammatory state favoring GCs at the expense of MCCs, there is an altered balance between Wnt and Notch signaling, in favor of Wnt signaling [30]. *WNT5A* has been associated with remodeling of airway smooth muscle cells in the context of airway hyperresponsiveness [58]. *WNT4* has been shown to be upregulated in the epithelium of patients suffering from chronic obstructive pulmonary disease and upregulates *IL8* and *CXCL8* gene expression in HAECs [59]. Interestingly, we found both *WNT5A* and *WNT4* specifically expressed by the subpopulation of SCs that may be involved in immune response. This finding again suggests a role for this SC population in the inflammation-induced airway remodeling.

Based on expression of the target genes *TCF4* and *SNAI2*, in our study, activation of the Wnt pathway appears to be confined to the BC population. This population also expresses strongly and specifically the ligand *WNT10A*, suggesting an autocrine regulatory loop. *WNT10A* has been shown to be specific of BC in other epithelia, such as the mammary epithelium [60]. In fallopian organoids, Wnt has been shown to be essential for stemness [61] and for self-renewal, but not proliferation, in basal-like breast cancer cells [62]. Thus, autocrine *WNT10A* signaling may regulate self-renewal in the BC compartment of the airway epithelium. In MCCs, we have observed a specific expression of both members from the Reptin family which are ATP-dependent DNA helicases that have been identified as Wnt signaling repressors [63,64]. As Reynolds and colleagues recently showed that  $\beta$ -catenin was necessary for mucus and ciliated cell specification [57], additional investigations should certainly be carried out to characterize precisely the role of Wnt/ $\beta$ -catenin during airway epithelial regeneration.

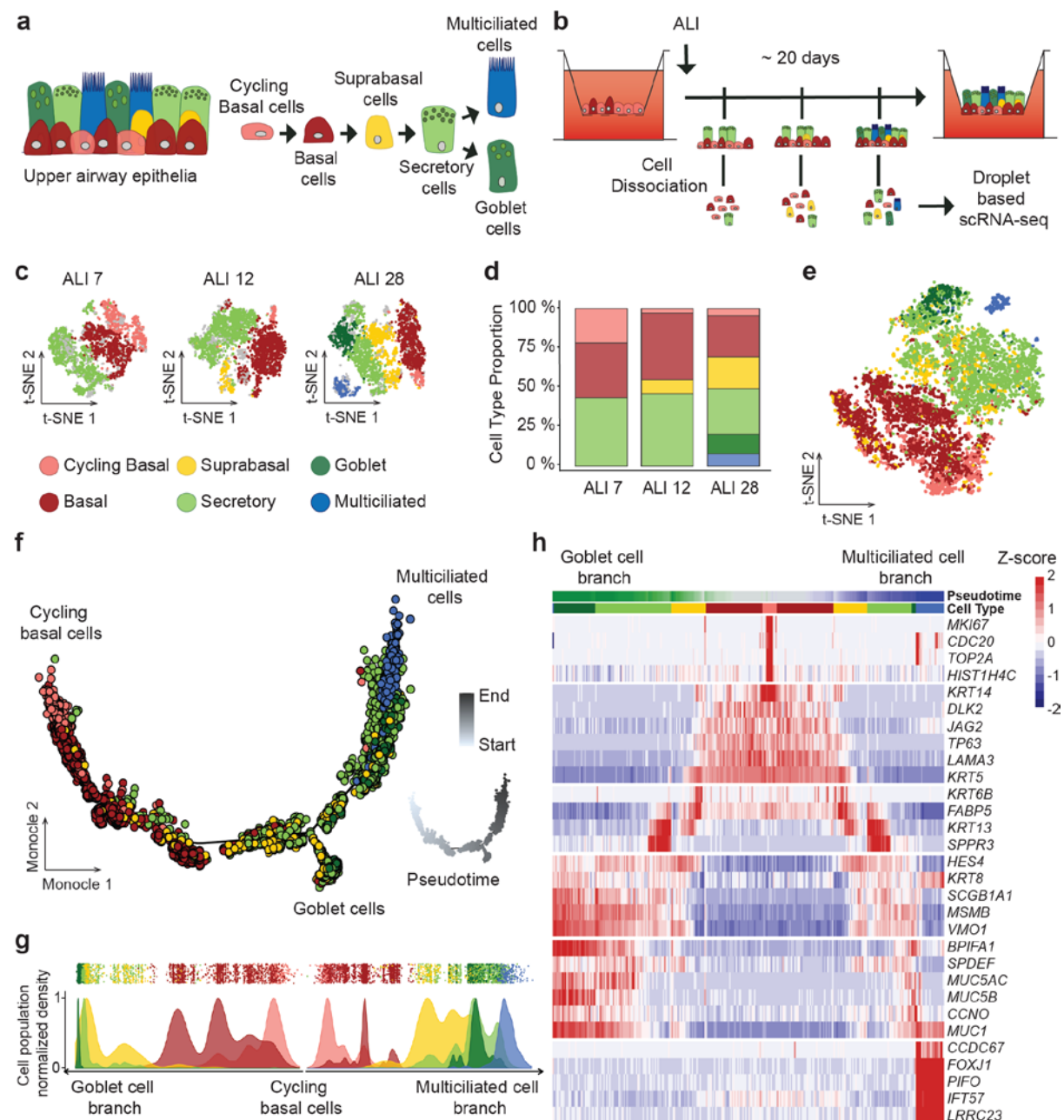
## **Conclusions**

Our lineage inference study during the airway epithelium regeneration has provided novel insights in differentiation dynamics by positioning goblet cells as possible precursors for multiciliated cells. Thus, in the airway epithelium, unlike in most tissues, cells carrying specialized function, i.e. secretory and goblet cells, can constitute differentiation intermediates for other specialized cells, the multiciliated cells. We have also identified subpopulations within the basal, suprabasal, secretory and multiciliated cell compartments. In particular, we propose that if all secretory cells produce anti-microbial peptides, one subset of secretory cells may be more specifically engaged in immune-cell signaling. Our dataset also provides extensive characterization of the deuterosomal cell population. In addition, we have established an exhaustive repertoire of keratin expression and showed that observation of the “keratin switch” during differentiation could be self-sufficient to establish the identity of the different cell



populations. Finally, we have improved signaling pathway characterization by detecting putative Notch repressors that might participate in Notch signal shutdown at the deuterosomal stage, and reporting Wnt pathway activity within the basal cell compartment.

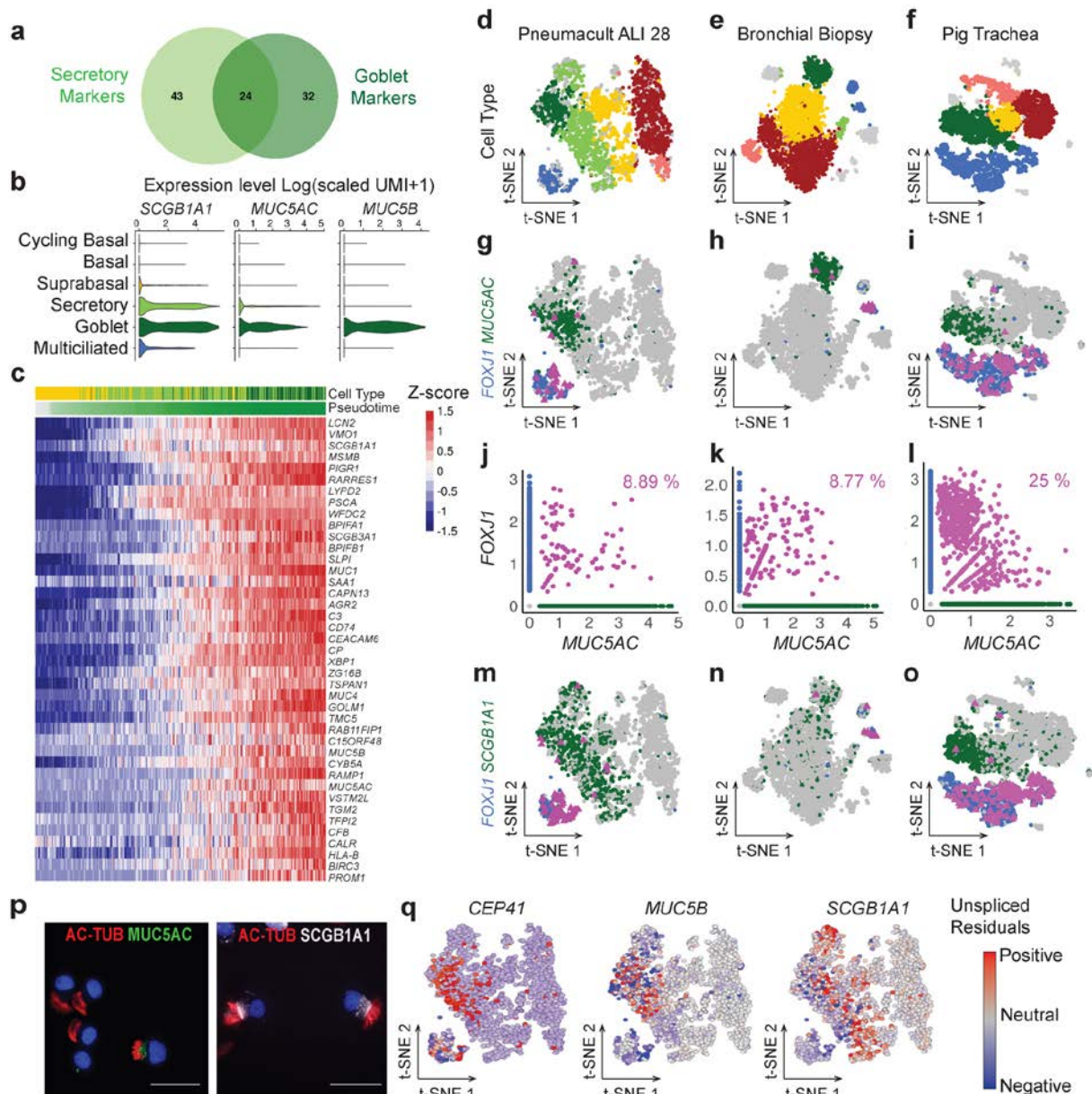
## Figures



**Figure 1: Characterization of MCC and Goblet cell lineages during airway epithelium regeneration using single cell RNA-Seq**

**a** Model of upper airway epithelium, based on 6 major types of epithelial cells, with consensus lineage hierarchy. **b** single-cell (sc)RNA-seq experimental design. Regenerating airway epithelia was dissociated at successive days (7, 12, 28) after a transition to an air-liquid interface (ALI). **c** t-SNE plots of the scRNA-seq expression data highlighting the main cell types observed at ALI7 (3426 cells), ALI12 (2785 cells), ALI28 (3615 cells) (grey: unassigned cells). **d** Relative abundance of the 6 main cell types at each time point. **e** Aggregate t-SNE plot of gene expression in 9826 cells. **f** Inference of goblet and MCC cell lineages by Monocle 2, based on an aggregate of the entire experiment. Color code is the

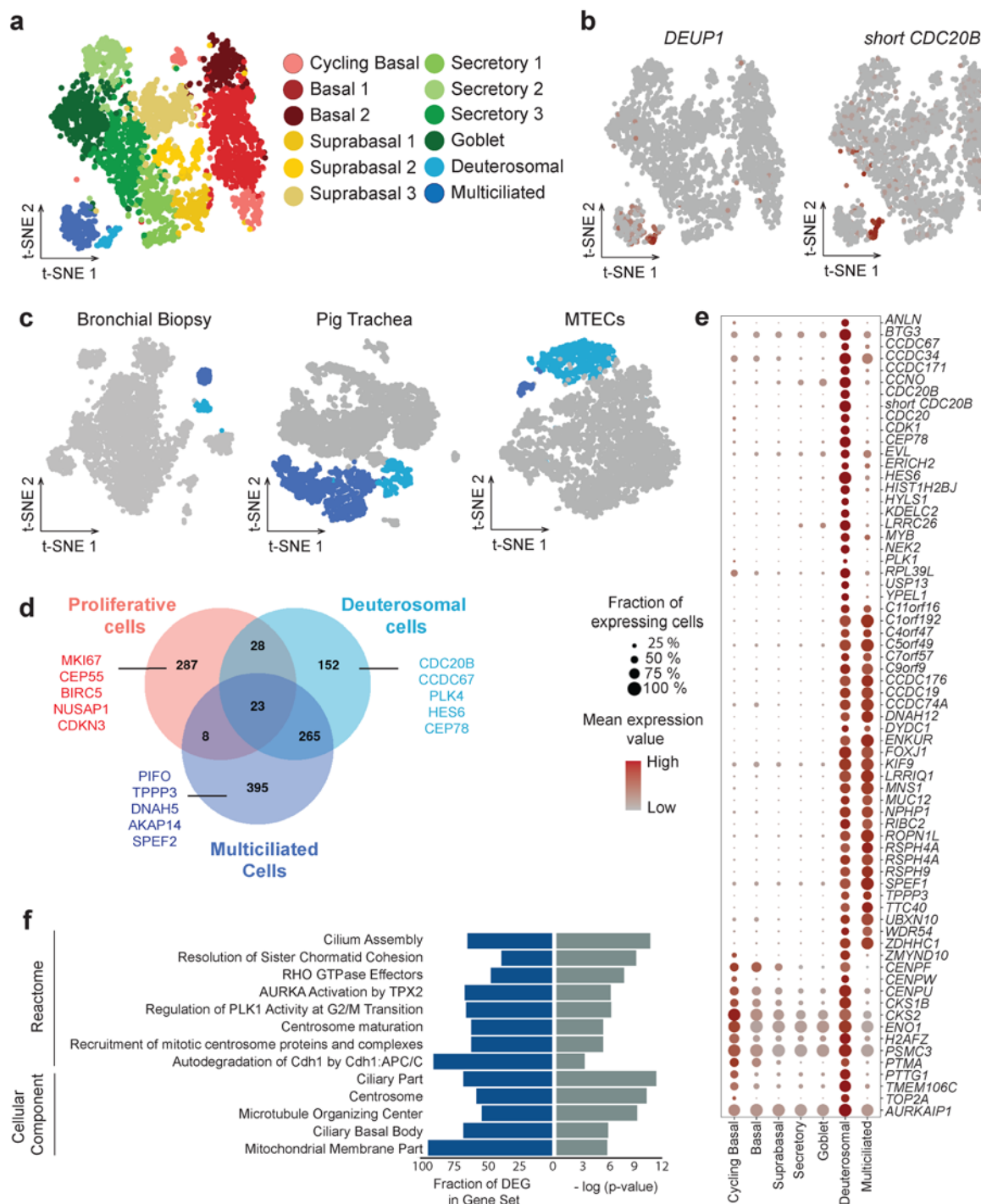
same as in c. Inset: pseudotime picturing by a white to grey gradient along the differentiation trajectory. **g** Distribution of the 6 main cell types in the pseudotime along the two branches of the trajectory (down: goblet cell branch; upper right: multiciliated cell branch). **h** Heatmap representing the smoothened temporal expression pattern of a representative list of cell type specific markers, with branch representation as in g. Cells were ordered by branch, then cluster appearance, then pseudotime.



**Figure 2: Goblet cells as differentiation intermediates for multiciliated cells**

**a** Venn diagram illustrating closeness of the best marker genes for secretory and goblet cells deduced from scRNA-seq of cells differentiated in Pneumacult medium (ALI28). **b** Violin plots of normalized expression for *SCGB1A1*, *MUC5AC* and *MUC5B*, three markers of secretory and goblet cells. **c** Heat map of the most differentially expressed genes between groups of suprabasal, secretory and goblet cells at key point in the pseudotime (before branching, start of the GC branch and end of the GC branch). Cells are ordered by pseudotime. Bars on the top of the heatmap indicate cell type and pseudotime. **d-e-f** t-SNE plots of expression from scRNA-seq of ALI28 (d), bronchial biopsy cells (e), pig tracheal cells (f). Colors indicates cell types as in Figure 1. **g-h-i** Highlights of gene expression for *FOXJ1*<sup>+</sup> cells (blue), *MUC5AC*<sup>+</sup> cells (green), and *FOXJ1*<sup>+</sup>/*MUC5AC*<sup>+</sup> cells (pink) in the same samples as in d-e-f. **j-k-l** Relationships between normalized expression of *MUC5AC* and *FOXJ1* in the three same samples. **m-n-**

o Highlights of gene expressions for *FOXJ1*+ cells (blue), *SCGB1A1*+ cells (green), and *FOXJ1*+/*SCGB1A1*+ cells (pink). p Immunodetection of cells co-expressing markers of multiciliated cells (Acetylated Tubulin) and of goblet cells (*MUC5AC*) (o) or of secretory cells (*SCGB1A1*) (p). q Representation by a t-SNE plot (scRNA-seq of cells differentiated in Pneumacult medium at ALI28) of the Velocity residuals colored according to estimates of the positive (red) and negative (blue) residues for a multiciliated cell marker (*CEP41*), a goblet cell marker (*MUC5B*) and a secretory cell marker (*SCGB1A1*).

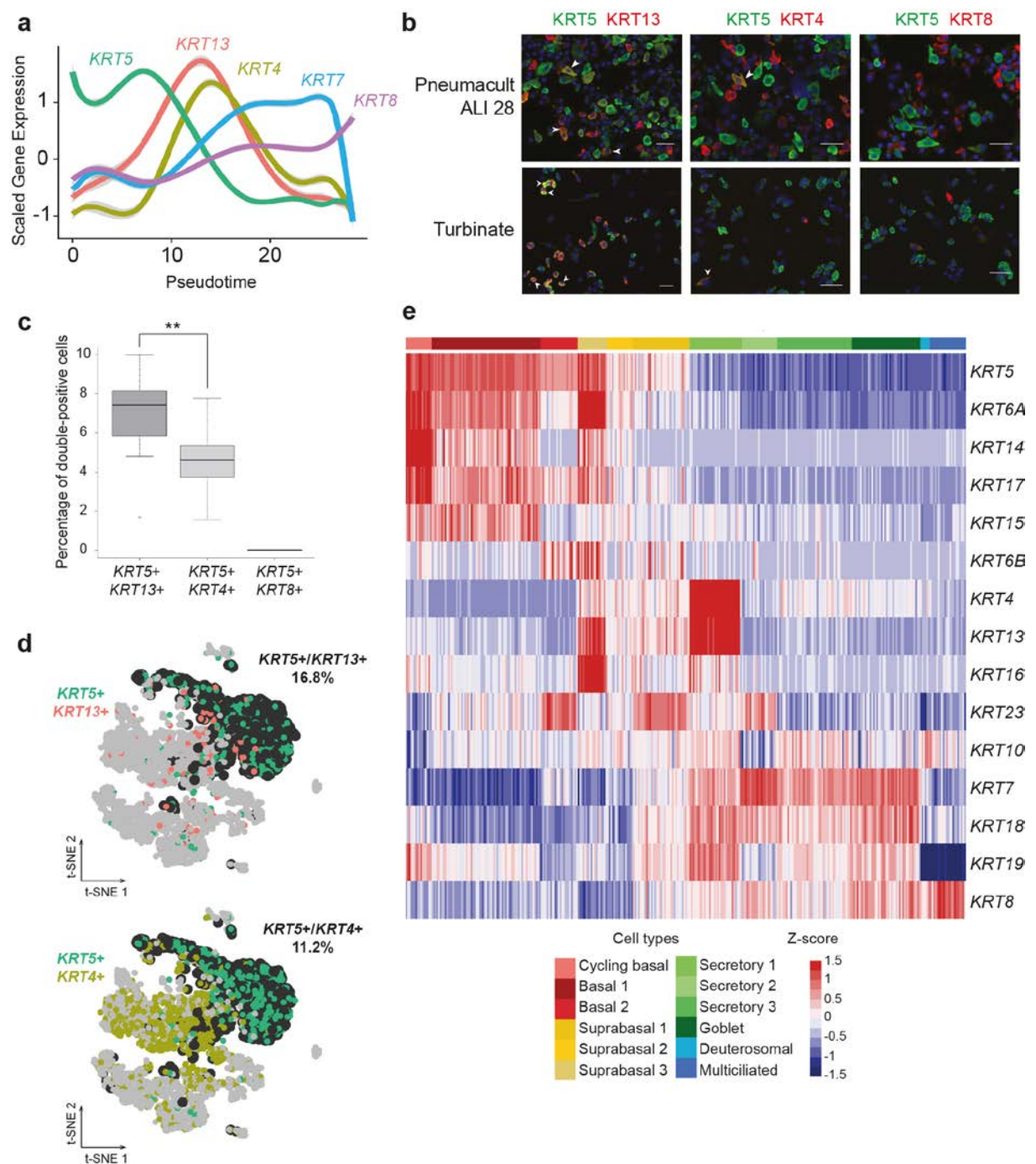


**Figure 3: Deuterosomal cells form a discrete MCC intermediate population with a centriole amplification signature**

**a** Subcluterization of scRNA-seq from cells differentiated in Pneumacult medium (ALI28) into 12 cell types, deduced from intra-heterogeneity analysis of the 6 initial clusters. **b** Illustration of the specific expression of *DEUP1* and *short CDC20B* in the deuterosomal cell population (low to high expression, grey to red). **c** Identification of the cluster of deuterosomal cells in scRNA-seq data from a biopsy of human bronchi, pig trachea and mouse primary culture (MTEC, ALI3). Deuterosomal (light blue). Multiciliated cluster (dark blue). **d** Venn Diagram showing overlaps existing between top gene markers

of deuterosomal cells (light blue) and those of proliferative (pink) or multiciliated cells (dark blue). **e** Dot plot of marker genes for the deuterosomal cell population. Color gradient (grey to red) and dot size indicate for each cluster the mean marker expression and the percentage of cells expressing the marker, respectively. **f** Enriched gene sets in deuterosomal cell marker genes.





**Figure 4: Keratin signatures switch during the differentiation process**

**a** Plot of normalized gene expression of keratins according to pseudotime from scRNA-seq of cells differentiated in Pneumacult medium (ALI28). **b** Double immunofluorescence staining for KRT5 and KRT13, KRT5 and KRT4 or KRT8. White arrowheads indicate doubly labelled cells (KRT5+/KRT13+, KRT5+/KRT4+, KRT5+/KRT8+). Nuclei are shown in blue (DAPI). **c** Quantification of double-positive cells from b. \*\*: pvalue<0.01 (Wilcoxon test). **d** tSNEs of scRNA-seq data from pig tracheal epithelial cells. KRT5+ cells are shown in green, KRT13+ are shown in red, KRT4+ are shown in light green, double-positive cells are

shown in black. The indicated percentage corresponds to double-positive cells. **e** Heatmap for scRNA-seq data from Pneumacult ALI28 showing gene expression for keratins.



## **Methods**

### **Human Airway Epithelial Cells (HAECs) culture**

HAECs cultures were derived from nasal mucosa of inferior turbinates. After excision, nasal inferior turbinates were immediately immersed in Ca<sup>2+</sup>/Mg<sup>2+</sup>-free HBSS supplemented with 25 mM HEPES, 200 U/mL penicillin, 200 µg/mL streptomycin, 50 µg/mL gentamicin sulfate, and 2.5 µg/ml amphotericin B (all reagents from Gibco). After repeated washes with cold supplemented HBSS, tissues were digested with 0.1% Protease XIV from *Streptomyces Griseus* (Sigma-Aldrich) overnight at 4°C. After incubation, fetal calf serum (FCS) was added to a final concentration of 10%, and nasal epithelial cells were detached from the stroma by gentle agitation. Cell suspensions were further dissociated by trituration through a 21G-needle and then centrifuged at 150g for 5 min. The pellet was resuspended in supplemented HBSS containing 10% FCS and centrifuged again. The second cell pellet was then suspended in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) containing 10% FCS and cells were plated (20 000 cells per cm<sup>2</sup>) on 75 cm<sup>2</sup>-flasks coated with rat-tail collagen I (Sigma-Aldrich). Cells were incubated in a humidified atmosphere of 5% CO<sub>2</sub> at 37°C. Culture medium was replaced with Bronchial Epithelium Basal Medium (BEBM, Lonza) supplemented with BEGM SingleQuot™ Kit Supplements (Lonza) on the day after and was then changed every other day. After 4 to 5 days of culture, after reaching about 70% confluence, cells were detached with trypsin-EDTA 0.05% (Gibco) for 5 min and seeded on Transwell® permeable supports (6.5 mm diameter; 0.4 µm pore size; Corning), in BEGM medium, with a density of 30 000 cells per Transwell®. Once the cells have reached confluence (typically after 5 days), they were induced to differentiate at the air-liquid interface by removing medium at the apical side of the Transwell®, and by replacing medium at the basal side with either DMEM:BEBM (1:1) supplemented with BEGM SingleQuot™ Kit Supplements or with Pneumacult-ALI (StemCell Technologies) as indicated in the figure legends. Culture medium was changed every other day.

### **Mouse tracheal epithelial cells (MTECs)**

MTECs cell cultures were established from the tracheas of 12 weeks-old mice. After dissection, tracheas were placed in cold DMEM:F-12 medium (1:1) supplemented with 15 mM HEPES, 100 U/mL penicillin, 100 µg/mL streptomycin, 50 µg/mL gentamicin sulfate, and 2.5 µg/ml amphotericin B. Each trachea was processed under a binocular microscope to remove as much conjunctive tissue as possible with small forceps and was opened longitudinally with small dissecting scissors. Tracheas were then placed in supplemented DMEM:F-12 containing 0.15% protease XIV from *Streptomyces Griseus*. After overnight incubation at 4°C, FCS was added to a final concentration of 10%, and tracheal epithelial cells were detached by gentle agitation. Cells were centrifuged at 400g for 10 min and resuspended in supplemented DMEM:F-12 containing 10% FCS. Cells were plated on regular cell culture plates and maintained in a humidified atmosphere of 5% CO<sub>2</sub> at 37°C for 4 hours to allow attachment of putative contaminating fibroblast. Medium containing cells in suspension was further centrifuged at 400g for 5 min and cells were resuspended in supplemented DMEM:F-12 containing BEGM SingleQuot™ kit supplements and 5% FCS. Cells were plated on rat tail collagen I-coated Transwell®. Typically, 5 tracheas resulted in 12



Transwells®. Medium was changed every other day. Air-liquid interface culture was conducted once transepithelial electrical resistance had reached a minimum of 1000 ohm/cm<sup>2</sup> (measured with EVOM2, World Precision Instruments). Air-liquid interface culture was obtained by removing medium at the apical side of the Transwell®, and by replacing medium at the basal side with Pneumacult-ALI medium (StemCell Technologies).

#### **HAEC and MTEC dissociation for single-cell RNA-seq**

Single-cell analysis was performed at the indicated days of culture at the air-liquid interface. To obtain a single-cell suspension, cells were incubated with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in supplemented HBSS for 4 hours at 4°C degrees. Cells were gently detached from Transwells® by pipetting and then transferred to a microtube. 50 units of DNase I (EN0523 Thermo Fisher Scientific) per 250 µL were directly added and cells were further incubated at room temperature for 10 min. Cells were centrifuged (150g for 5 min) and resuspended in 500 µL supplemented HBSS containing 10% FCS, centrifuged again (150g for 5 min) and resuspended in 500 µL HBSS before being mechanically dissociated through a 26G syringe (4 times). Finally, cell suspensions were filtered through a 40 µm porosity Flowmi™ Cell Strainer (Bel-Art), centrifuged (150 g for 5 min) and resuspended in 500 µL of cold HBSS. Cell concentration measurements were performed with Scepter™ 2.0 Cell Counter (Millipore) and Countess™ automated cell counter (Thermo Fisher Scientific). Cell viability was checked with Countess™ automated cell counter (Thermo Fisher Scientific). All steps except the DNase I incubation were performed on ice. For the cell capture by the 10X genomics device, the cell concentration was adjusted to 300 cells/µl in HBSS aiming to capture 1500 cells for HAECs and 5000 cells for MTECs.

#### **Turbinate epithelial cell dissociation**

To obtain a single-cell suspension from turbinates, the whole turbinate was incubated with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in supplemented HBSS at 4°C degrees overnight. Epithelial cells were gently detached from the turbinate by washing with HBSS pipetting up and down and then transferred to a 50 ml Falcon tube. Cells were centrifuged (150g for 5 min at 4°C), after removing of the supernatant the cells were resuspended in 1 ml of HBSS, 50 units of DNase I (EN0523 Thermo Fisher Scientific) per 250 µL were directly added and cells were further incubated at room temperature for 10 min. Cells were centrifuged (150g for 5 min at 4°C) and resuspended in 1 ml supplemented HBSS containing 10% FCS, centrifuged again (150g for 5 min at 4°C) and resuspended in 500 µL HBSS before being mechanically dissociated through a 26G syringe (4 times). Finally, cell suspensions were filtered 40 µm porosity Flowmi™ Cell Strainer (Bel-Art), centrifuged (150 g for 5 min) and resuspended in 500 µL of cold HBSS. Cell concentration measurements were performed with Scepter™ 2.0 Cell Counter (Millipore) and Countess™ automated cell counter (Thermo Fisher Scientific). Cell viability was checked with Countess™ automated cell counter (Thermo Fisher Scientific). All steps except the DNase I incubation were performed on ice. For the cell capture by the 10X genomics device, the cell concentration was adjusted to 500 cells/µl in HBSS aiming to capture 5000 cells.

### **Anesthetic procedure**

Intranasal anesthesia is performed with topical application (gauze) of 5% lidocaine (anesthetic) plus naphazoline (vasoconstrictor) solution (0.2 mg/ml). Laryngeal and endobronchial anesthesia is performed with topical application of 2% lidocaine through the working channel of a 4.9 mm outer diameter bronchoscope.

### **Nasal brushing**

Brushing was performed with a 2 mm cytology brush (Medi-Globe) in the inferior turbinate zone.

### **Bronchial biopsy**

Bronchial biopsy was performed at the spur between the left upper lobe and the left lower lobe with a 1.8mm-diameter Flexibite biopsy forceps (Medi-Globe) passed through the working channel of the bronchoscope (WCB).

### **Dissociation of nasal brushing**

The brush was soaked in a 5 mL Eppendorf containing 1 mL of dissociation buffer which was composed of HypoThermosol® (BioLife Solutions) 10 mg/mL protease from *Bacillus Licheniformis* (Sigma-Aldrich, reference P5380) and 0.5 mM EDTA. The tube was shaken vigorously and centrifuged for 2 min at 150 g. The brush was removed, cells pipetted up and down 5 times and then incubated cells on ice for 30 min, with gentle trituration with 21G needles 5 times every 5 min. Protease was inactivated by adding 200 µL of HBSS/2% BSA. Cells were centrifuged (400g for 5 min at 4°C). Supernatant was discarded leaving 10 µL of residual liquid on the pellet. Cells were resuspended in 500 µL of wash buffer (HBSS/0.05% BSA) and 2.250 mL of Ammonium Chloride 0.8% was added to perform red blood cell lysis. After a 10 min incubation, 2 mL of wash buffer were added, and cells were centrifuged (400g for 5 min at 4°C). Supernatant was discarded leaving 10 µL of residual liquid on the pellet, cells were resuspended in 1000 µL of wash buffer centrifuged (400g for 5 min at 4°C). Supernatant was discarded leaving 10 µL of residual liquid on the pellet, cells were resuspended in 1000 µL of wash buffer and passed through 40 µm porosity Flowmi™ Cell Strainer (Bel-Art), then centrifuged (400g for 5 min at 4°C). Supernatant was discarded leaving 10 µL of residual liquid on the pellet, cells were resuspended in 100 µL of wash buffer. Cell counts and viability were performed with Countess™ automated cell counter (Thermo Fisher Scientific). For the cell capture by the 10X genomics device, the cell concentration was adjusted to 500 cells/µl in HBSS aiming to capture 5000 cells. All steps were performed on ice.

### **Dissociation of bronchial biopsy**

The biopsy was soaked in 1 mL dissociation buffer which was composed of DPBS, 10 mg/mL protease from *Bacillus Licheniformis* (Sigma-Aldrich, reference P5380) and 0.5 mM EDTA. After 1 h, the biopsy was finely minced with a scalpel, and returned to dissociation buffer. From this point, the dissociation procedure is the same as the one described in the “dissociation of nasal brushing” section, with an incubation time increased to 1h, and

omitting the red blood cell lysis procedure. For the cell capture by the 10X genomics device, the cell concentration was adjusted to 300 cells/ $\mu$ L in HBSS aiming to capture 5000 cells. All steps were performed on ice.

### **Pig tracheal epithelial cell dissociation**

To obtain a single-cell suspension from pig trachea, whole clean tracheas were incubated with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in supplemented HBSS at 4°C degrees overnight. Epithelial cells were gently detached from the turbinate by washing with HBSS pipetting up and down and then transferred to a 50 ml Falcon tube. Cells were centrifuged (150g for 5 min at 4°C), after removing of the supernatant the cells were resuspended in 1 ml of HBSS, 50 units of DNase I (EN0523 Thermo Fisher Scientific) per 250  $\mu$ L were directly added and cells were further incubated at room temperature for 10 min. Cells were centrifuged (150g for 5 min at 4°C) and resuspended in 1 ml supplemented HBSS containing 10% FCS, centrifuged again (150g for 5 min at 4°C) and resuspended in 500  $\mu$ L HBSS before being mechanically dissociated through a 26G syringe (4 times). Finally, cell suspensions were filtered through 40  $\mu$ m porosity Flowmi™ Cell Strainer (Bel-Art), centrifuged (150 g for 5 min) and resuspended in 500  $\mu$ L of cold HBSS. Cell concentration measurements were performed with Scepter™ 2.0 Cell Counter (Millipore) and Countess™ automated cell counter (Thermo Fisher Scientific). Cell viability was checked with Countess™ automated cell counter (Thermo Fisher Scientific). All steps except the DNase I incubation were performed on ice. For the cell capture by the 10X genomics device, the cell concentration was adjusted to 500 cells/ $\mu$ L in HBSS aiming to capture 5000 cells.

### **Single-cell RNA-seq**

We followed the manufacturer's protocol (Chromium™ Single Cell 3' Reagent Kit, v2 Chemistry) to obtain single cell 3' libraries for Illumina sequencing. Libraries were sequenced with a NextSeq 500/550 High Output v2 kit (75 cycles) that allows up to 91 cycles of paired-end sequencing: Read 1 had a length of 26 bases that included the cell barcode and the UMI; Read 2 had a length of 57 bases that contained the cDNA insert; Index reads for sample index of 8 bases. Cell Ranger Single-Cell Software Suite v1.3 was used to perform sample demultiplexing, barcode processing and single-cell 3' gene counting using standards default parameters and human build hg19, pig build sus scrofa 11.1 and mouse build mm10.

### **Single-cell quantitative PCR**

HAECs were dissociated as described above, then single cells were separated with a C1™ Single-cell AutoPrep system (Fluidigm), followed by quantitative PCR on the Biomark system (Fluidigm) using SsoFast™ evaGreen® Supermix (Biorad) and the primers described in Supplementary Table S2.

### **RNA-seq on dissociated and non-dissociated HAECs**



Two Transwells® from fully differentiated HAECs from 2 distinct donors were each dissociated as described above. After the final resuspension, cells were centrifuged and resuspended in 800 µL Qiazol (Qiagen). Non-dissociated cells from 2 Transwells® were also lysed in 800 µL Qiazol. RNAs were extracted with the miRNeasy mini kit (Qiagen) according to manufacturer's instructions. Two micrograms from each RNA was used in RNA-seq library construction with the Truseq® stranded total RNA kit (Illumina). Sequencing was performed with a NextSeq 500/550 High Output v2 kit (75 cycles). Reads were aligned against hg19 human build using STAR aligner. Low expressed genes were filtered out, then paired differential analysis was performed with DESeq2 comparing dissociated vs non-dissociated samples from cultures generated from 2 different donors. P-values were adjusted for multiple testing using the false discovery rate (FDR). Top differentially expressed genes were selected using the following cutoffs: FDR < 0.001 and an absolute log2FC > 1.5

### Cytospins

Fully differentiated HAECs were dissociated by incubation with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in HBSS (Hanks' balanced salts) overnight at 4°C. Cells were gently detached from the Transwells® by pipetting and then transferred to a microtube. Cells were then cytocentrifuged at 72 g for 10 min onto SuperFrost™ Plus slides using a Shandon Cytospin™ 4 cytocentrifuge.

### Tissue processing and OCT tissue embedding (see if we maintain figures containing this part)

Nasal turbinates were fixed in PFA 4% at 4°C overnight then, for cryoprotection, tissues were soaked in a 15% sucrose solution until tissue sinks, then soaked in a 30% solution before tissue sinks. Tissue was embedded in "optimal cutting temperature" OCT medium (Thermo Fisher Scientific) at room temperature and then submerged in Isopentane previously tempered at -80°C. Cutting of frozen tissues was performed with a cryostat Leica CM3050 S.

### Immunostaining

Cytospin™ slides were fixed for 10 min in 4% paraformaldehyde at room temperature and then permeabilized with 0.5% Triton X-100 in PBS for 10 min. Cells were blocked with 3% BSA in PBS for 30 min. The incubation with primary antibodies was carried out at 4°C overnight.

Primary antibodies: mouse monoclonal KRT4 (1:50, Santa Cruz Biotechnology sc-52321), mouse monoclonal KRT8 (1:50, Santa Cruz Biotechnology, sc-58737), mouse monoclonal KRT13 (1:200, Sigma-Aldrich clone KS-1A3), rabbit KRT5 (1:2000, Biolegend BLE905501), Rabbit CC10 (SCGB1A1) (1:500, Millipore 07-623), mouse monoclonal Acetylated Tubulin (1:500, Sigma-Aldrich clone 6-11B-1), mouse monoclonal MUC5AC (1:250, Abnova clone 45M1).

Secondary antibodies: Alexa Fluor 488 goat anti-rabbit (1:500; Thermo Fisher Scientific), Alexa Fluor 647 goat anti-mouse (1:500; Thermo Fisher Scientific), Alexa Fluor 488 Goat anti-Mouse IgG1 (1:500, Fisher Scientific), Alexa Fluor 594 Goat anti-Mouse IgG2a (1:500, Fisher Scientific), Alexa Fluor 647 Goat anti-Mouse IgG2b (1:500,

Fisher Scientific). Incubation with secondary antibodies was carried out during 1h at room temperature. Nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI).

When necessary, Acetylated Tubulin, Muc5AC and KRT5 antibodies were directly coupled to CF™ 594/488/488 respectively, with the Mix-n-Stain™ kit (Sigma-Aldrich) according to the manufacturer's instruction. Coupled primary antibodies were applied for 2 hours at room temperature after secondary antibodies had been extensively washed and after a 30 min blocking stage in 3% normal rabbit or mouse serum in PBS. MTECs immunostainings were directly performed on Transwell® membranes using a similar protocol. For mounting on slides, Transwell® membranes were cut with a razor blade and mounted with ProLong™ Gold medium (Thermo Fisher Scientific). Images were acquired using the Olympus Fv10i or Leica sp5 confocal imaging systems.

### **Time course sample analysis**

*Preprocessing.* For each sample, cells with levels in the top 5% or bottom 5% of distribution for the following quality metrics: number of expressed features, dropout percentage and library size (total UMI count) were filtered out. Additionally, cells with a percentage of mitochondrial genes > top 5% were also removed. Quality metrics were computed using the scater package (2.3.0) [65]. Genes expressed at less than 1 UMI in at least 5 cells were removed from further analysis.

*Normalization.* The scran package [66] was used to calculate cell-based scale factors and normalize cells for differences in count distribution. Each sample was normalized separately twice, first in an unsupervised manner, then after grouping cells of similar gene expression based on our robust clustering results.

*Clustering Robustness.* In order to best determine the key steps in the studied differentiation process, a customized method was implemented to analyze clustering robustness to dataset perturbation. For all possible number of clusters (from 2 to 9), multiple subsets of the studied dataset were created (10 subsets, with 10% of the cells randomly removed each time) and clustering was performed multiple times on each subset with changing settings of the seed parameter. The result of those clusterings were stored in a n cells x n cells stability matrix, containing for each pair of cells 1 or 0 if the cells are clustered together or not. This stability matrix was then transformed in a Euclidean distance matrix between cells and then divided into the used k number of clusters k using hierarchical clustering (hclust with 'average' method). To identify the optimal number of clusters, a visual inspection of the elbow plot of the average intra-stability (mean stability within each cluster) and the average inter-stability (mean stability between each cluster) was done. Cells with a stability metric less than 70% were labelled as Unassigned, due to their high clustering variability between each round of clustering, then removed from further analysis of the time course data. Cell clustering was performed using SIMLR (package version 1.4.1) [67].

*Differential analysis.* To further analyze the robustness of each step of the differentiation process, we tested the robustness of the cell type marker gene identification through differential gene expression analysis. Differential expression analysis was performed using edgeR (package version 3.22) [68]. In a one vs. all differential analysis, a pool of 100 cells from one cluster were analyzed against an equal mixture of cells from all other clusters. In a one vs one differential analysis pools of cells of the same size were compared. Those differential analysis were

performed multiple times (10 times) on different pool of cells and the DEG identified were compared between each pool of cells using the rank-rank hypergeometric overlap algorithm [69]. Unfortunately, this approach was too stringent, and only identified highly expressed marker gene as they are less likely to submit to dropout events. Thus, the Seurat FindAllMarkers function based on a non-parametric Wilcoxon rank sum test was used to identify cell type marker genes.

*Time points aggregation.* 10X datasets generated during the time course were aggregated using MNN correction [70] from the scran package.

*Trajectory Inference.* Trajectory inference was performed using monocle 2 (package version 2.8) [71]. Cell ordering was based on highly variable genes (~ 200 - 500 genes) selected by their expression dispersion. Monocle analysis on the aggregated time points was done on raw count after library size correction (downsampling). Branch building was performed using BEAM analysis within the monocle package, and corresponding differential analysis was done by cross comparison of group of cells along the pseudotime (before branching, after branching and at the branch end) using Seurat 1 vs 1 differential analysis.

*Cell type projection.* To compare cell types identified in distinct samples, cells were projected from one dataset onto the other using scmap R package version 1.1, scmapCluster function [72].

*Data Visualization.* All graphs were generated using R (ggplot2). Heatmaps were obtained using pheatmap (no clustering used, genes ordered by their expression in pseudotime or in cluster, cells ordered by pseudotime or cluster). Heatmaps show smoothed gene expression values: for each gene, normalized gene expression values were first transformed into z-scores, then averaged across 10 neighboring cells in the chosen ordering (pseudotime only or pseudotime in clusters). For single gene representation: only cells with expression levels above the top 50 percentiles for that gene are represented for clarity.

## Individual sample analysis

Each sample of our study was reanalyzed with less stringent parameters, to identify rare or transitory cell types or gene expression events

*Preprocessing, normalization and clustering.* Individual dataset analysis was performed using Seurat standard analysis pipeline. Briefly, cells were first filtered based on number of expressed features, dropout percentage, library size and mitochondrial gene percentage. Thresholds were selected by visually inspecting violin plots in order to remove the most extreme outliers. Genes expressing less than 5 UMI across all cells were removed from further analysis. Cell-level normalization was performed using the median UMI counts as scaling factor. Highly variable genes were selected for following analysis based on their expression level and variance. PCA analysis was performed on those hvgs, the number of PCs to use was chosen upon visual inspection of the PC variance elbowplot (~10 to 20 PCs depending on the dataset). Clustering was first performed with default parameter and then increasing the resolution parameter above 0.5 to identify small clusters (but with the knowledgeable risk of

splitting big cluster due to high gene expression variability). Differential analysis was again performed using Seurat FindAllMarkers and FindMarkers functions based on non-parametric Wilcoxon rank sum test. Gene Set Enrichment analysis was performed using fgsea R package with the following gene sets reactome.db (R package) and GO cellular component (Broad Institute GSEA MSigDB) genesets. Molecular function enrichment analysis was performed using Ingenuity Pathway Analysis (Qiagen).

*Cell Type Annotation.* Based on the time course experiment analysis and associated top ~15 marker genes identified, a score was computed to associate cell types to each cluster. Scoring method is based on Macosko. *et al.* cell cycle phase assignment [73]. For each cell it measures the mean expression of the top marker genes for each possible cell type, which results in a matrix c cell types per n cells. Then it performs a z-score of those mean expression for each cell, the top resulting score gives the matching cell type.

*Velocity.* RNA velocity was calculate using latest release of velocityto pipeline (<http://velocityto.org/>) using standard parameters: GTF file used for Cell Ranger analysis and the possorted\_genome\_bam.bam, Cell Ranger output alignment file. From the loom file which contains count table of spliced and unspliced transcript the gene.relative.velocity.estimates function was used on cell type marker genes. The resulting expression pattern of unspliced-spliced phase portraits shows the induction or repression of those marker genes from one cell type to the next. We used velocityto package version 0.5 [33].

#### **Plasscheart *et al.* dataset**

Plasscheart *et al.* data [29] were downloaded as processed data along with visualization coordinates and were used without further manipulation.

([https://kleintools.hms.harvard.edu/tools/springViewer\\_1\\_6\\_dev.html?datasets/reference\\_HBECs/reference\\_HBECs](https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?datasets/reference_HBECs/reference_HBECs))

## **Declarations**

### **Ethics approval and consent to participate**

Human samples were collected according to the guidelines of the Declaration of Helsinki, after approval by the institutional review board "Comité de Protection des Personnes Sud Est IV " (12/12/2017). All patients gave their written informed consent to participate. Mouse experiments were performed following the Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. The IPMC is authorized to use animals by the French agricultural ministry (Ministere de l'Agriculture et de l'Alimentation) and persons performing experiments were authorized by the French research ministry (Ministère de l'Education Supérieure, de la Recherche et de l'Innovation).

### **Consent for publication**

All patients gave their written informed consent for anonymous publication of data arising from their samples.

### **Availability of data and materials**

All datasets generated during this study were deposited to the Gene Expression Omnibus repository under the series number GSE121600.

### **Competing interests**

The authors declare no competing financial interests.

### **Funding**

This project was funded by grants from FRM (DEQ20180339158), the labex Signallife (ANR-11-LABX-0028-01), the association Vaincre la Mucoviscidose (RF20180502280), and the Chan Zuckerberg Initiative (Silicon Valley Foundation, 2017-175159 -5022). The UCAGenomiX platform, a partner of the National Infrastructure France Génomique, is supported by Commissariat aux Grands Investissements (ANR-10-INBS-09-03, ANR-10-INBS-09-02) and Canceropôle PACA.

### **Authors' contributions**

LEZ and PB designed and supervised the study and obtained fundings. SRG, MJA, AC, VM and LEZ performed the experiments. SL et CHM provided human fresh tissues. IC provided pig trachea. MD, KLB and AP performed the bioinformatics analysis. All authors were involved in data interpretation. MD and LEZ designed the figures. LEZ drafted the original manuscript. MD, PB, SRG, AP, KLB, BM and LEZ edited the manuscript.

## **Acknowledgements**

We are grateful to the UCAGenomiX platform and to Marin Truchi for fruitful discussions and technical help on single cell RNA sequencing.

## References

1. Gras D, Chanez P, Vachier I, Petit A, Bourdin A. Bronchial epithelium as a target for innovative treatments in asthma. *Pharmacol Ther* [Internet]. Elsevier Inc.; 2013;140:290–305. Available from: <http://dx.doi.org/10.1016/j.pharmthera.2013.07.008>
2. Kotton DN, Morrissey EE. Lung regeneration: mechanisms, applications and emerging stem cell populations. *Nat Med* [Internet]. 2014;20:822–32. Available from: <http://dx.doi.org/10.1038/nm.3642>
3. Cohn L. Mucus in chronic airway diseases: Sorting out the sticky details. *J Clin Invest*. 2006;116:306–8.
4. Merigo F, Benati D, Piacentini G, Boner A, Sbarbati A. The ultrastructure of nasal mucosa in children with asthma. *Ultrastruct Pathol*. 2002;26:293–8.
5. Curran DR, Cohn L. Advances in mucous cell metaplasia: A plug for mucus as a therapeutic focus in chronic airway disease. *Am J Respir Cell Mol Biol*. 2010;42:268–75.
6. Rock JR, Onaitis MW, Rawlins EL, Lu Y, Clark CP, Xue Y, et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A* [Internet]. 2009;106:12771–5. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2714281&tool=pmcentrez&rendertype=abstract>
7. Cole BB, Smith RW, Jenkins KM, Graham BB, Reynolds PR, Reynolds SD. Tracheal basal cells: A facultative progenitor cell pool. *Am J Pathol* [Internet]. American Society for Investigative Pathology; 2010;177:362–76. Available from: <http://dx.doi.org/10.2353/ajpath.2010.090870>
8. Hogan BLM, Barkauskas CE, Chapman HA, Epstein JA, Jain R, Hsia CCW, et al. Repair and regeneration of the respiratory system: Complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* [Internet]. Elsevier Inc.; 2014;15:123–38. Available from: <http://dx.doi.org/10.1016/j.stem.2014.07.012>
9. Boers JE, Ambergen AW, Thunnissen FB. Number and Proliferation of Basal and Parabasal Cells in Normal Human Airway Epithelium. *Am J Respir Crit Care Med* [Internet]. 1998;157:2000–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10228131> <http://www.atsjournals.org/doi/pdf/10.1164/10.1164/ajrccm.157.6.9707011>



10. Pardo-Saganta A, Law BMM, Tata PRR, Villoria J, Saez B, Mou H, et al. Injury Induces Direct Lineage Segregation of Functionally Distinct Airway Basal Stem/Progenitor Cell Subpopulations. *Cell Stem Cell* [Internet]. Elsevier Inc.; 2015;16:184–97. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S193459091500003X>
11. Watson JKK, Rulands S, Wilkinson ACC, Wuidart A, Ousset M, Van Keymeulen A, et al. Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium. *Cell Rep* [Internet]. 2015;12:90–101. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S2211124715006099>
12. Boers JE, Ambergen AW, Thunnissen FBJM. Number and Proliferation of Clara Cells in Normal Human Airway Epithelium. *Am J Respir Crit Care Med*. 1999;159.
13. Wang S-Z, Rosenberger CL, Bao Y-X, Stark JM, Harrod KS. Clara Cell Secretory Protein Modulates Lung Inflammatory and Immune Responses to Respiratory Syncytial Virus Infection. *J Immunol* [Internet]. 2003;171:1051–60. Available from: <http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.171.2.1051>
14. Jones KG, Holland JF, Foureman GL, Bend JR, Fouts JR. Xenobiotic metabolism in Clara cells and alveolar type II cells isolated from lungs of rats treated with beta-naphthoflavone. *J Pharmacol Exp Ther* [Internet]. 1983 [cited 2018 Jun 18];225:316–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6601704>
15. Rawlins EL, Okubo T, Xue Y, Brass DM, Auten RL, Hasegawa H, et al. The Role of Scgb1a1+ Clara Cells in the Long-Term Maintenance and Repair of Lung Airway, but Not Alveolar, Epithelium. *Cell Stem Cell* [Internet]. Elsevier Ltd; 2009;4:525–34. Available from: <http://dx.doi.org/10.1016/j.stem.2009.04.002>
16. Chen G, Korfhagen TR, Xu Y, Kitzmiller J, Wert SE, Maeda Y, et al. SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production. *J Clin Invest*. 2009;119:2914–24.
17. Morimoto M, Liu Z, Cheng HT, Winters N, Bader D, Kopan R. Canonical Notch signaling in the developing lung is required for determination of arterial smooth muscle cells and selection of Clara versus ciliated cell fate. *J Cell Sci* [Internet]. 2010/01/06. 2010;123:213–24. Available from: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=20048339](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20048339)
18. Tsao P-NN, Vasconcelos M, Izvolsky KI, Qian J, Lu J, Cardoso W V. Notch signaling controls the

balance of ciliated and secretory cell fates in developing airways. *Development* [Internet].

2009/06/09. The Company of Biologists Limited; 2009 [cited 2015 Jun 12];136:2297–307. Available from:

[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=19502490](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19502490)

19. Pardo-Saganta A, Tata PR, Law BM, Saez B, Chow RD-W, Prabhu M, et al. Parent stem cells can serve as niches for their daughter cells. *Nature* [Internet]. 2015;523:597–601. Available from:

<http://www.nature.com/doifinder/10.1038/nature14553>

20. Rock JR, Gao X, Xue Y, Randell SH, Kong Y-Y, Hogan BLM. Notch-Dependent Differentiation of Adult Airway Basal Stem Cells. *Cell Stem Cell* [Internet]. 2011 [cited 2015 Apr 18];8:639–48. Available from: <http://www.sciencedirect.com/science/article/pii/S1934590911001688>

21. Marcet B, Chevalier B, Luxardi G, Coraux C, Zaragosi L-EE, Cibois M, et al. Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nat Cell Biol* [Internet]. 2011/05/24. Nature Publishing Group; 2011;13:693–9. Available from:

[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=21602795](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21602795)

22. Marcet B, Chevalier B, Coraux C, Kodjabachian L, Barbry P. MicroRNA-based silencing of Delta/Notch signaling promotes multiple cilia formation. *Cell Cycle* [Internet]. 2011/08/23. 2011;10:2858–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21857154>

23. Fletcher RB, Das D, Gadye L, Street KN, Baudhuin A, Wagner A, et al. Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell* [Internet]. Elsevier Inc.; 2017;20:817–830.e8. Available from: <http://dx.doi.org/10.1016/j.stem.2017.04.003>

24. Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol* [Internet]. 2018 [cited 2018 Jun 20];19:85–97. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29167569>

25. Pal B, Chen Y, Vaillant F, Jamieson P, Gordon L, Rios AC, et al. Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat Commun* [Internet]. Nature Publishing Group; 2017 [cited 2018 Jun 20];8:1627. Available from:

<http://www.nature.com/articles/s41467-017-01560-x>

26. Plass M, Solana J, Alexander Wolf F, Ayoub S, Misios A, Glažar P, et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* (80- ). 2018;360.

27. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* [Internet]. Nature Publishing Group; 2014 [cited 2014 Apr 28];509:371–5. Available from: <http://www.nature.com/gate1.inist.fr/nature/journal/v509/n7500/full/nature13173.html>
28. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket S, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* [Internet]. Springer US; 2018;560:319–24. Available from: <http://dx.doi.org/10.1038/s41586-018-0393-7>
29. Plasschaert LW, Žilionis R, Choo-wing R, Savova V, Knehr J, Roma G, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature*. 2018;560:377–381.
30. Ordovas-Montanes J, Dwyer DF, Nyquist SK, Buchheit KM, Vukovic M, Deb C, et al. Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* [Internet]. Springer US; 2018; Available from: <http://www.nature.com/articles/s41586-018-0449-8>
31. Chevalier B, Adamiok A, Mercey O, Revinski DR, Zaragosi L-E, Pasini A, et al. MiR-34/449 control apical actin network formation during multiciliogenesis through small GTPase pathways. *Nat Commun*. 2015;6.
32. Revinski DR, Zaragosi L-E, Boutin C, Pons N, Marcet B, Kodjabachian L, et al. CDC20B is required for deuterosome-mediated centriole production in multiciliated cells. *Nat Commun*. 2018;In Press.
33. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature* [Internet]. Springer US; 2018;560:494–8. Available from: <http://www.nature.com/articles/s41586-018-0414-6>
34. Zaragosi L, Revinski DR, Mercey O, Chevalier B, Kodjabachian L, Marcet B, et al. CDC20B, the host gene of miR-449, controls vertebrate multiciliogenesis. *Prep*. 2017;
35. Al Jord A, Shihavuddin A, Servignat d’Aout R, Faucourt M, Genovesio A, Karaïskou A, et al. Calibrated mitotic oscillator drives motile ciliogenesis. *Science* (80- ) [Internet]. 2017 [cited 2018 Jan 2];358:803–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28982797>
36. Colopy SA, Bjorling DE, Mulligan WA, Bushman W. A population of progenitor cells in the basal and intermediate layers of the murine bladder urothelium contributes to urothelial development and regeneration. *Dev Dyn* [Internet]. 2014 [cited 2018 Sep 21];243:988–98. Available from: <http://doi.wiley.com/10.1002/dvdy.24143>
37. Hudson DL, Guy AT, Fry P, O’Hare MJ, Watt FM, Masters JRW. *Epithelial Cell Differentiation*

Pathways in the Human Prostate: Identification of Intermediate Phenotypes by Keratin Expression. *J Histochem Cytochem* [Internet]. 2001 [cited 2018 Sep 21];49:271–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11156695>

38. Jumppanen M, Gruvberger-Saal S, Kauraniemi P, Tanner M, Bendahl P-O, Lundin M, et al. Basal-like phenotype is not associated with patient survival in estrogen-receptor-negative breast cancers. *Breast Cancer Res* [Internet]. 2007 [cited 2018 Sep 21];9:R16. Available from: <http://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr1649>

39. Yang L, Nichols J, Yao C, Maniley J, Robey E, Weinmaster G. Fringe Glycosyltransferases Differentially Modulate Notch1 Proteolysis Induced by Delta1 and Jagged1. *Mol Biol Cell* [Internet]. 2004;16:927–42. Available from: <http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E04-07-0614>

40. Bae S, Bessho Y, Hojo M, Kageyama R. The bHLH gene Hes6, an inhibitor of Hes1, promotes neuronal differentiation. *Development* [Internet]. 2000;127:2933–43. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10851137>

41. Cruciat CM. Casein kinase 1 and Wnt/ $\beta$ -catenin signaling. *Curr Opin Cell Biol* [Internet]. Elsevier Ltd; 2014;31:46–55. Available from: <http://dx.doi.org/10.1016/j.ceb.2014.08.003>

42. Tözser J, Earwood R, Kato A, Brown J, Tanaka K, Didier R, et al. TGF- $\beta$  Signaling Regulates the Differentiation of Motile Cilia. *Cell Rep*. 2015;11:1000–7.

43. Chand HS, Harris JF, Tesfaigzi Y. IL-13 in LPS-Induced Inflammation Causes Bcl-2 Expression to Sustain Hyperplastic Mucous cells. *Sci Rep* [Internet]. Nature Publishing Group; 2018 [cited 2018 Oct 13];8:436. Available from: <http://www.nature.com/articles/s41598-017-18884-9>

44. Tyner JW, Kim EY, Ide K, Pelletier MR, Roswit WT, Morton JD, et al. Blocking airway mucous cell metaplasia by inhibiting EGFR antiapoptosis and IL-13 transdifferentiation signals. *J Clin Invest*. 2006;116:309–21.

45. Turner J, Roger J, Fitau J, Combe D, Giddings J, Van Heeke G, et al. Goblet cells are derived from a FOXJ1-expressing progenitor in a human airway epithelium. *Am J Respir Cell Mol Biol*. 2011;44:276–84.

46. Gomperts BN, Kim LJ, Flaherty SA, Hackett BP. IL-13 regulates cilia loss and foxj1 expression in human airway epithelium. *Am J Respir Cell Mol Biol*. 2007;37:339–46.

47. Pardo-Saganta A, Law BM, Gonzalez-Celeiro M, Vinarsky V, Rajagopal J. Ciliated cells of pseudostratified airway epithelium do not become mucous cells after ovalbumin challenge. *Am J*

Respir Cell Mol Biol. 2013;48:364–73.

48. Reynolds SD, Reynolds PR, Pryhuber GS, Finder JD, Stripp BR. Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways. *Am J Respir Crit Care Med*. 2002;166:1498–509.

49. Guha A, Vasconcelos M, Zhao R, Gower AC, Rajagopal J, Cardoso W V. Analysis of notch signaling-dependent gene expression in developing airways reveals diversity of clara cells. *PLoS One*. 2014;9.

50. Mori M, Mahoney JE, Stupnikov MR, Paez-Cortez JR, Szymaniak AD, Varelas X, et al. Notch3-Jagged signaling controls the pool of undifferentiated airway progenitors. *Development* [Internet]. The Company of Biologists Limited; 2015 [cited 2015 Jun 15];142:258–67. Available from: <http://dev.biologists.org/content/142/2/258.full>

51. Wang X, Gupta P, Fairbanks J, Hansen D. Protein kinase CK2 both promotes robust proliferation and inhibits the proliferative fate in the *C. elegans* germ line. *Dev Biol* [Internet]. Elsevier; 2014;392:26–41. Available from: <http://dx.doi.org/10.1016/j.ydbio.2014.05.002>

52. Cheng P, Kumar V, Liu H, Youn J-I, Fishman M, Sherman S, et al. Effects of Notch Signaling on Regulation of Myeloid Cell Differentiation in Cancer. *Cancer Res* [Internet]. 2014 [cited 2018 Oct 15];74:141–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24220241>

53. Nam SM, Kim YN, Kim JW, Kyeong DS, Lee SH, Son Y, et al. Hairless and Enhancer of Split 6 (Hes6) Deficiency in Mouse Impairs Neuroblast Differentiation in Dentate Gyrus Without Affecting Cell Proliferation and Integration into Mature Neurons. *Cell Mol Neurobiol*. Springer US; 2016;36:57–67.

54. Fernandez-Martinez J, Vela EM, Tora-Ponsioen M, Ocaña OH, Nieto MA, Galceran J. Attenuation of Notch signalling by the Down-syndrome-associated kinase DYRK1A. *J Cell Sci* [Internet]. The Company of Biologists Ltd; 2009 [cited 2018 Oct 15];122:1574–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19383720>

55. Hsieh JJ, Zhou S, Chen L, Young DB, Hayward SD. CIR, a corepressor linking the DNA binding factor CBF1 to the histone deacetylase complex. *Proc Natl Acad Sci U S A* [Internet]. 1999;96:23–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9874765>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC15086>

56. Estrach S, Ambler CA, Lo Celso CL, Hozumi K, Watt FM. Jagged 1 is a -catenin target gene required for ectopic hair follicle formation in adult epidermis. *Development* [Internet].

2006;133:4427–38. Available from: <http://dev.biologists.org/cgi/doi/10.1242/dev.02644>

57. Malleske DT, Hayes D, Lallier SW, Hill CL, Reynolds SD. Regulation of human airway epithelial tissue stem cell differentiation by  $\beta$ -catenin, P300, and CBP. *Stem Cells* [Internet]. 2018 [cited 2018 Oct 15]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30171668>

58. Koopmans T, Kumawat K, Halayko AJ, Gosens R. Regulation of actin dynamics by WNT-5A: implications for human airway smooth muscle contraction. *Sci Rep* [Internet]. 2016 [cited 2018 Oct 15];6:30676. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27468699>

59. Durham AL, McLaren A, Hayes BP, Caramori G, Clayton CL, Barnes PJ, et al. Regulation of Wnt4 in chronic obstructive pulmonary disease. *FASEB J* [Internet]. 2013;27:2367–81. Available from: <http://www.fasebj.org/doi/10.1096/fj.12-217083>

60. Ji H, Goode RJAA, Vaillant F, Mathivanan S, Kapp EA, Mathias RA, et al. Proteomic profiling of secretome and adherent plasma membranes from distinct mammary epithelial cell subpopulations. *Proteomics* [Internet]. 2011 [cited 2018 Oct 15];11:4029–39. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21834135>

61. Kessler M, Hoffmann K, Brinkmann V, Thieck O, Jackisch S, Toelle B, et al. The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. *Nat Commun* [Internet]. Nature Publishing Group; 2015;6:1–11. Available from: <http://dx.doi.org/10.1038/ncomms9989>

62. DiMeo TA, Anderson K, Phadke P, Feng C, Perou CM, Naber S, et al. A Novel Lung Metastasis Signature Links Wnt Signaling with Cancer Cell Self-Renewal and Epithelial-Mesenchymal Transition in Basal-like Breast Cancer. *Cancer Res* [Internet]. 2009 [cited 2018 Oct 15];69:5364–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19549913>

63. Bauer A, Chauvet S, Huber O, Usseglio F, Rothbacher U, Aragnol D, et al. Pontin52 and Reptin52 function as antagonistic regulators of beta-catenin signalling activity. *EMBO J* [Internet]. 2000 [cited 2018 Oct 15];19:6121–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11080158>

64. Weiske J, Huber O. The histidine triad protein Hint1 interacts with Pontin and Reptin and inhibits TCF-  $\beta$ -catenin-mediated transcription. *J Cell Sci* [Internet]. 2005 [cited 2018 Oct 15];118:3117–29. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16014379>

65. McCarthy DJ, Campbell KR, Lun ATL, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* [Internet]. Oxford

University Press; 2017 [cited 2018 Oct 19];33:btw777. Available from:

<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw777>

66. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Research [Internet]. 2016 [cited 2018 Oct 19];5:2122.

Available from: <https://f1000research.com/articles/5-2122/v2>

67. Wang B, Zhu J, Pierson E, Ramazzotti D, Batzoglou S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. Nat Methods [Internet]. 2017 [cited 2018 Oct

19];14:414–6. Available from: <http://www.nature.com/articles/nmeth.4207>

68. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics [Internet]. 2010 [cited 2018 Oct 19];26:139–

40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19910308>

69. Plaisier SB, Taschereau R, Wong JA, Graeber TG. Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. Nucleic Acids Res [Internet]. 2010 [cited 2018 Oct 19];38:e169–e169. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/20660011>

70. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol [Internet]. 2018 [cited 2018 Oct

19];36:421–7. Available from: <http://www.nature.com/doifinder/10.1038/nbt.4091>

71. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14:979–82.

72. Kiselev VY, Yiu A, Hemberg M. scmap: projection of single-cell RNA-seq data across data sets. Nat Methods [Internet]. 2018 [cited 2018 Oct 19];15:359–62. Available from:

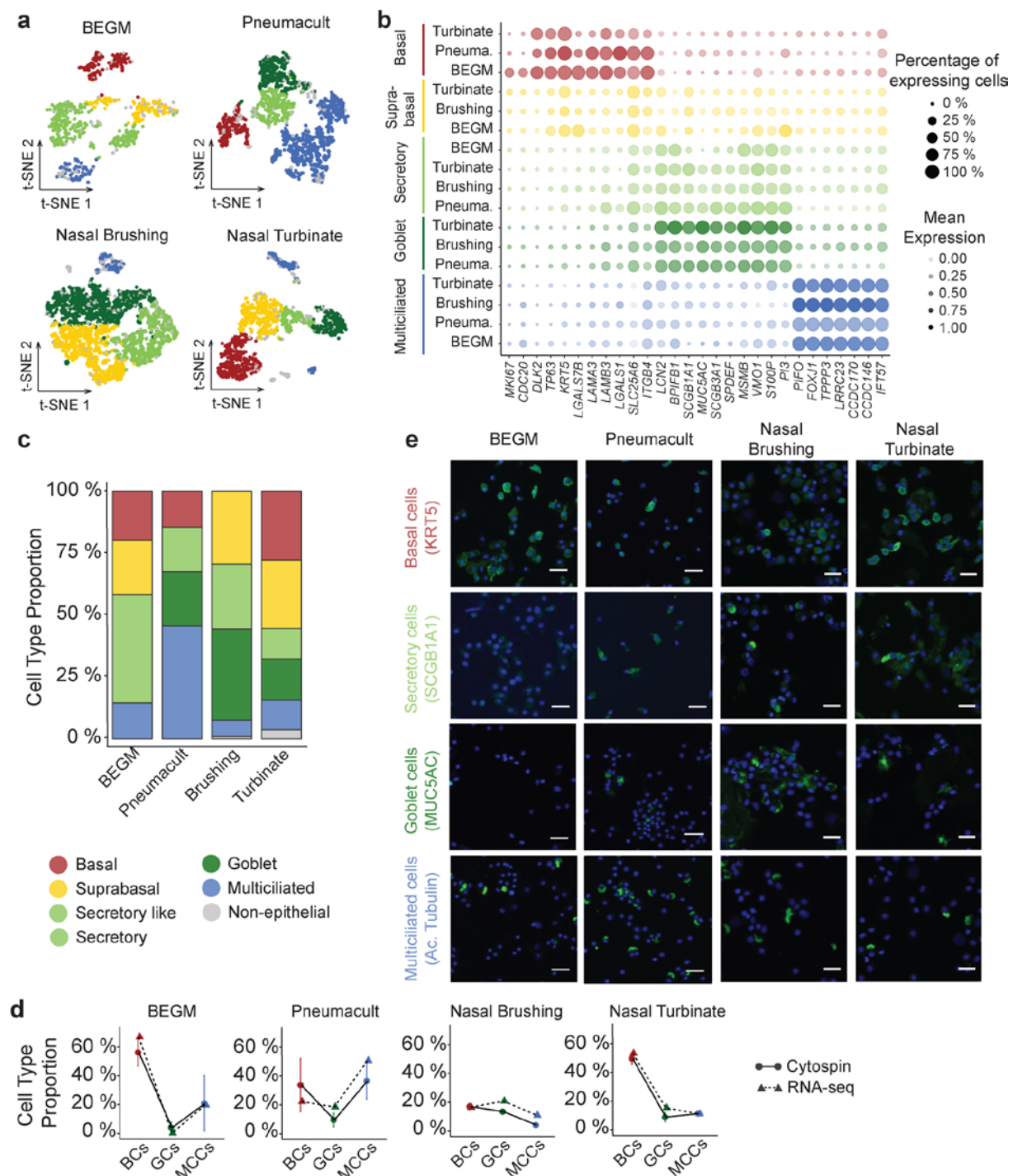
<http://www.ncbi.nlm.nih.gov/pubmed/29608555>

73. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell [Internet]. 2015 [cited 2018

Jan 25];161:1202–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26000488>



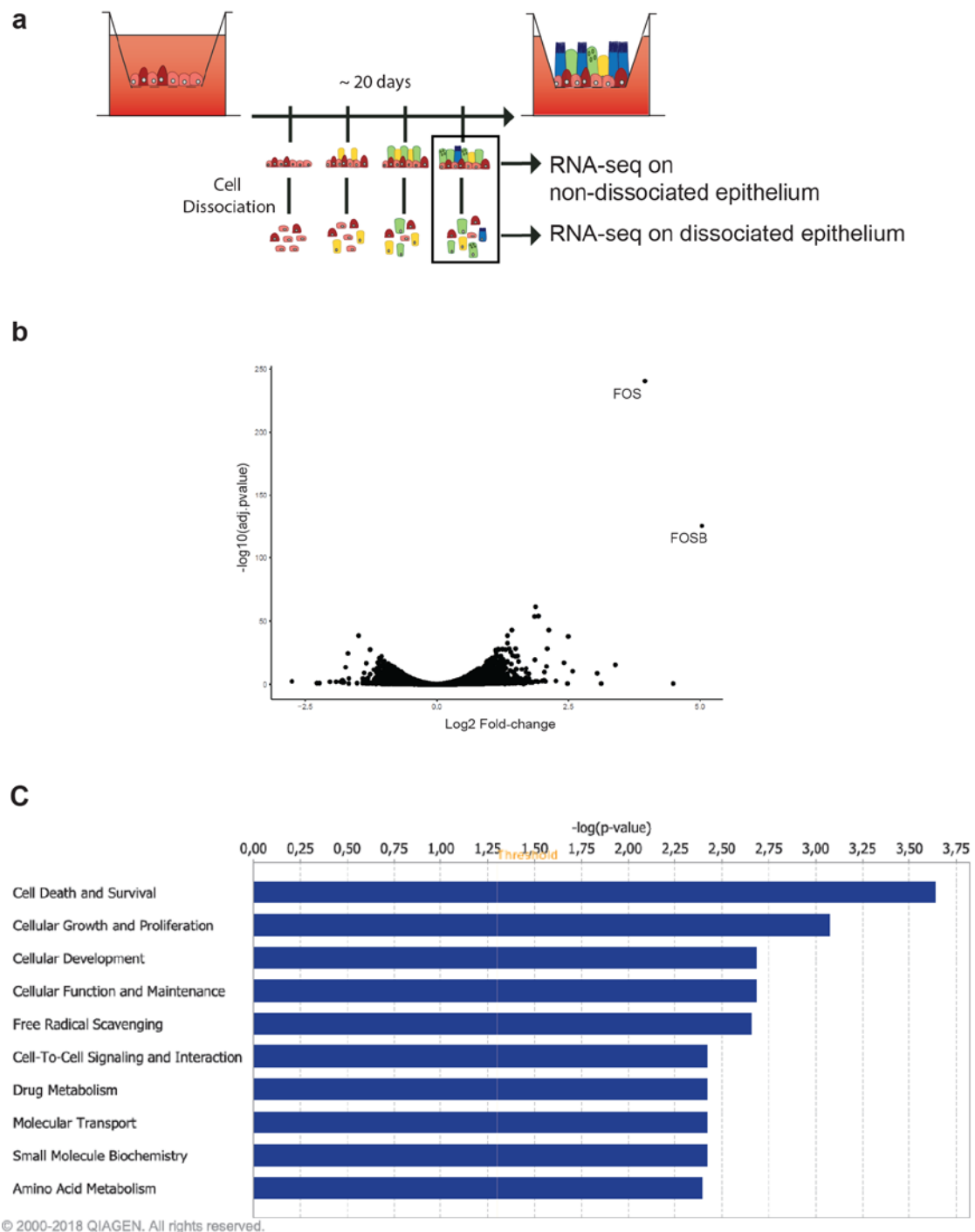
## Supplementary figures



**Supplementary figure 1: Cell type composition comparison between homeostatic *in vitro* samples and fresh human tissues**

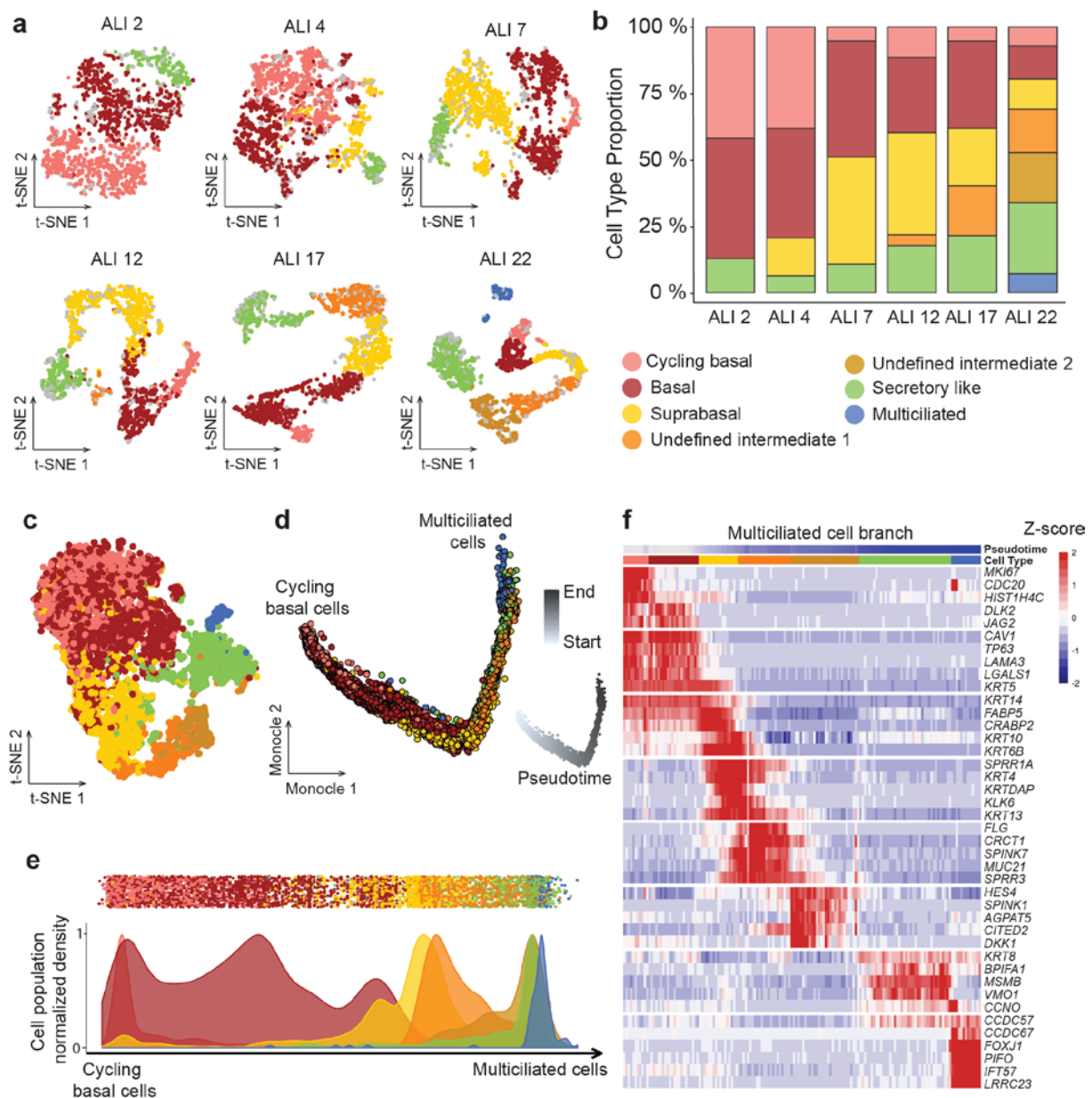
**a** t-SNE representing the distinct cell populations identified in each sample. **b** Dot plot of the main cell population marker genes. Dot size describe the percentage of cells expressing the respective marker genes and the average expression level of that gene based on UMI counts are shown by color intensity. **c** Relative abundance of cell types in each sample. **d** Scatter plot of the percentage of cells expressing

the selected marker genes for basal, goblet and multiciliated cells at the transcript (dotted line) and protein level (full line). **e** Immunostaining of the selected marker genes for basal, secretory, goblet and multiciliated in cytospin for each sample.



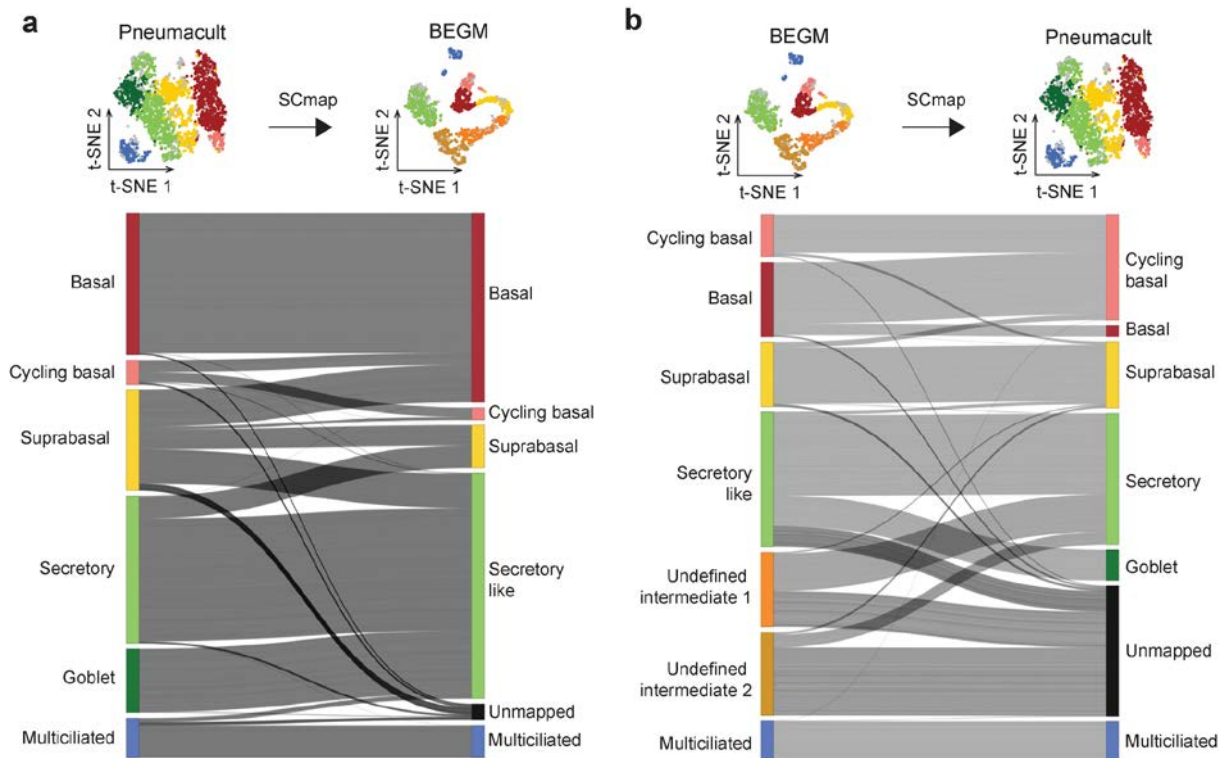
## Supplementary figure 2: Effect of cell dissociation on HAEC gene expression

**a** RNA-seq experimental design. Regenerating airway epithelia was dissociated at 28 days after a transition to an air-liquid interface. Dissociated and non-dissociated cultures from 2 donors were subjected to RNA-seq. **b** Volcano plot showing differential gene expression of dissociated vs. non-dissociated cell cultures. **c** Identification, with Ingenuity Pathway Analysis (Qiagen) of the molecular functions most significantly affected by dissociation, based on an analysis of 300 differential expressed genes ( $\text{FDR} < 0.01$  and  $\text{abs}(\log_2\text{FC}) > 1$ ).



**Supplementary figure 3: Characterization of MCC cell lineages during airway epithelium regeneration using single cell RNA-Seq in BEGM medium.**

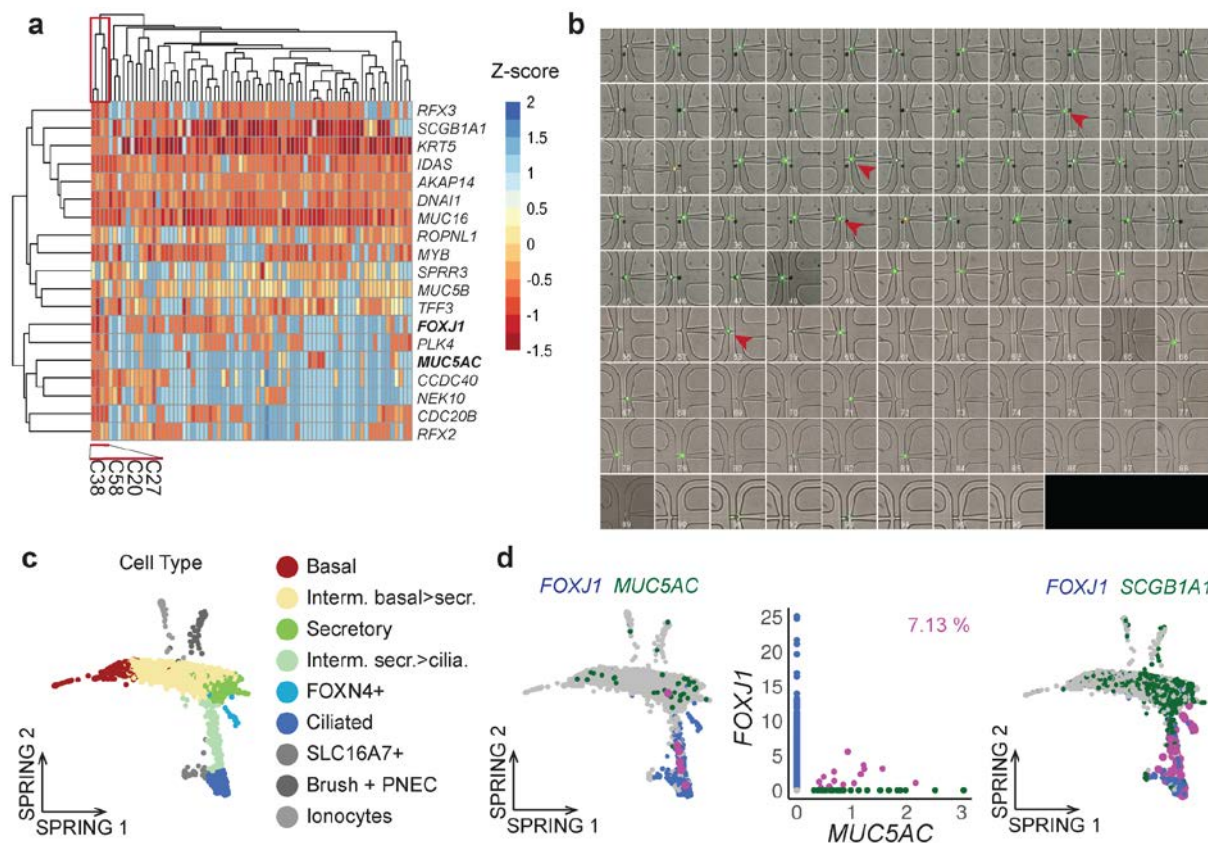
**a** t-SNE representing the distinct cell populations identified at each time points. **b** Relative abundance of cell types at each time point. **c** t-SNE plot of the aggregate of all cells from each time point. **d** Representation of the cell lineages inferred by Monocle 2 occurring during the upper airway epithelium regeneration (aggregate of all time points). Pseudotime evolution along the differentiation trajectory shown by white to grey gradient. **e** Distribution of the defined cell types in pseudotime. **f** Heatmap representing the smoothed temporal expression pattern of indicated cell type specific marker genes on MCC trajectory. Cells were ordered by cluster appearance in pseudotime.



**Supplementary figure 4: Comparison of fully differentiated epithelia cell population between Pneumacult and BEGM media.**

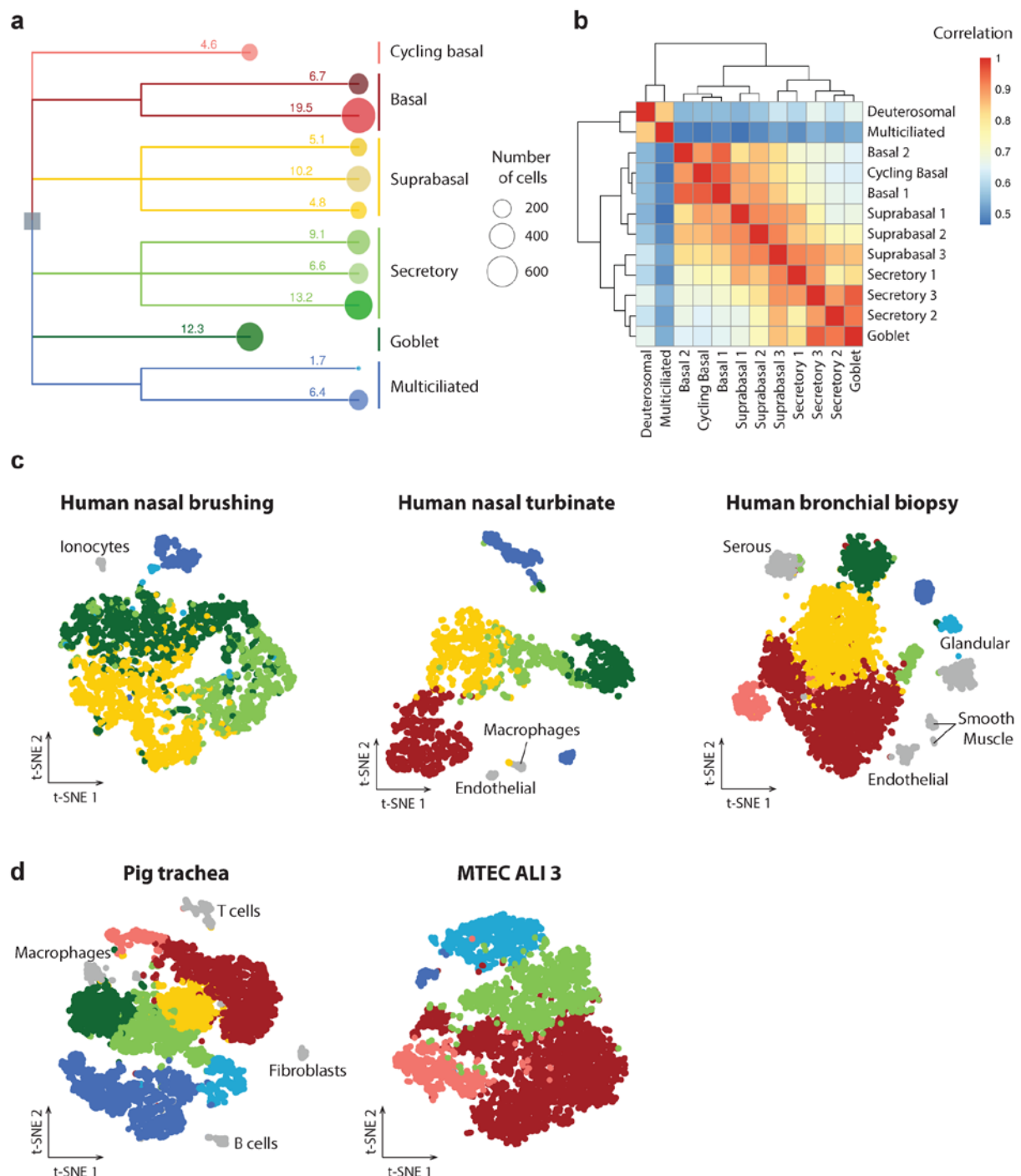
**a** Sankey Network of the mapping of Pneumacult cells onto BEGM cells. **b** Sankey Network of the mapping of BEGM cells onto Pneumacult cells.





**Supplementary figure 5: Validation of Goblet cells markers in multiciliated cells in doublet free single cell transcriptomic datasets.**

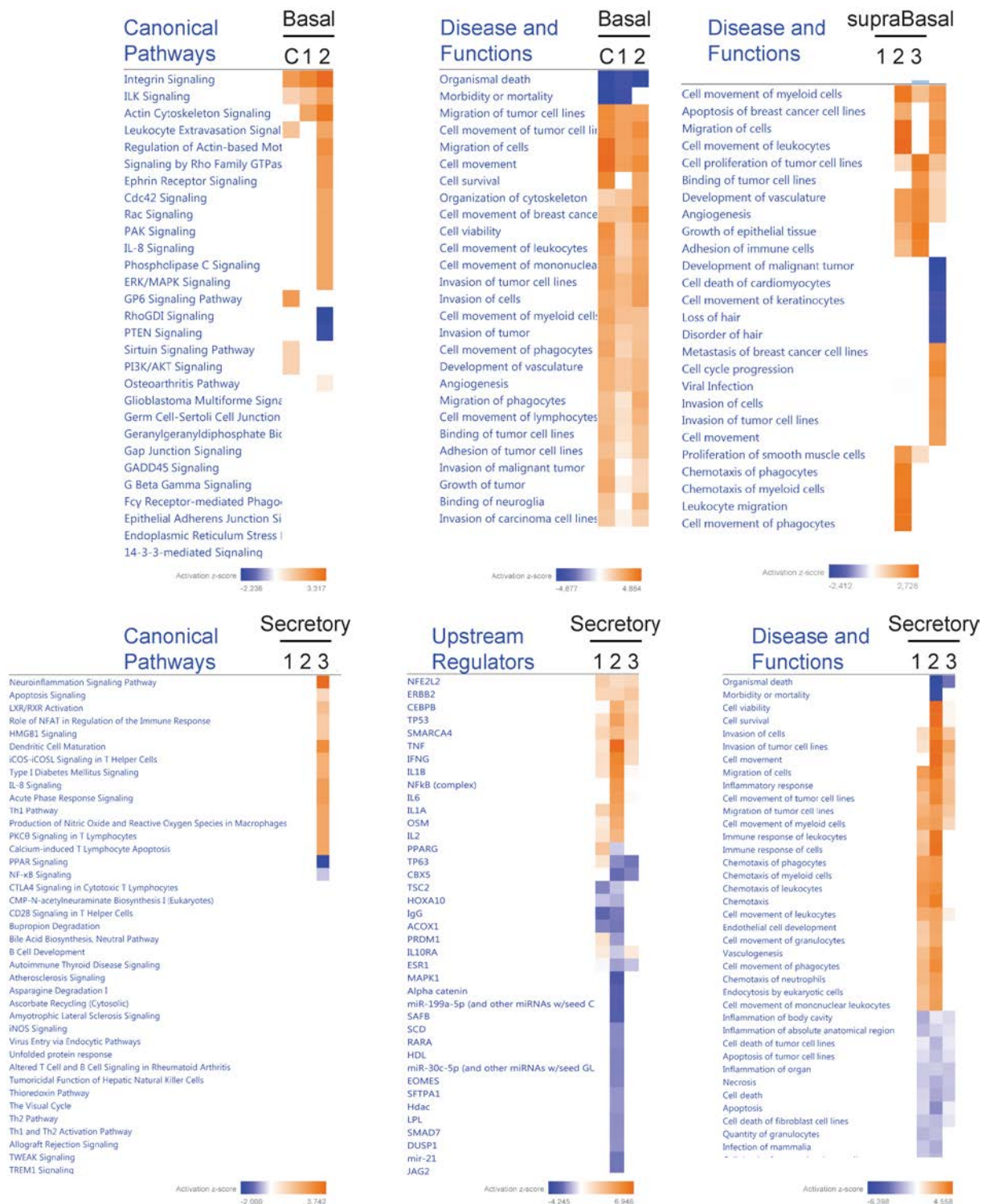
**a** Heatmap colored by Z-score from C1 Biomark experiment in PneumaCult media, co-expression of Goblet and Multiciliated cell markers in single cells (red cluster). **b** C1™ Single-Cell Preamp IFC (10–17  $\mu$ m) imaging, Red arrowheads show cells expressing both cell type markers (FOXJ1 and MUC5AC), one cell per chamber, green cells=living cells, red cells=dead cells. **c** SPRING representations of Plasschaert *et al.* dataset. Left panel colored by cell type. Center left panel colored by FOXJ1+ cells (blue), MUC5AC+ cells (green), co-expressing FOXJ1 and MUC5AC cells (pink). Center right panel displays a scatter plot of normalized expression of MUC5AC and FOXJ1 in cells (dot) colored by FOXJ1+ cells (blue), MUC5AC+ cells (green), co-expressing FOXJ1 and MUC5AC cells (pink). Right panel colored by FOXJ1+ cells (blue), SCGB1A1+ cells (green), co-expressing FOXJ1 and SCGB1A1 cells (pink).



### Supplementary figure 6: Identification of subpopulations by high resolution cell clustering

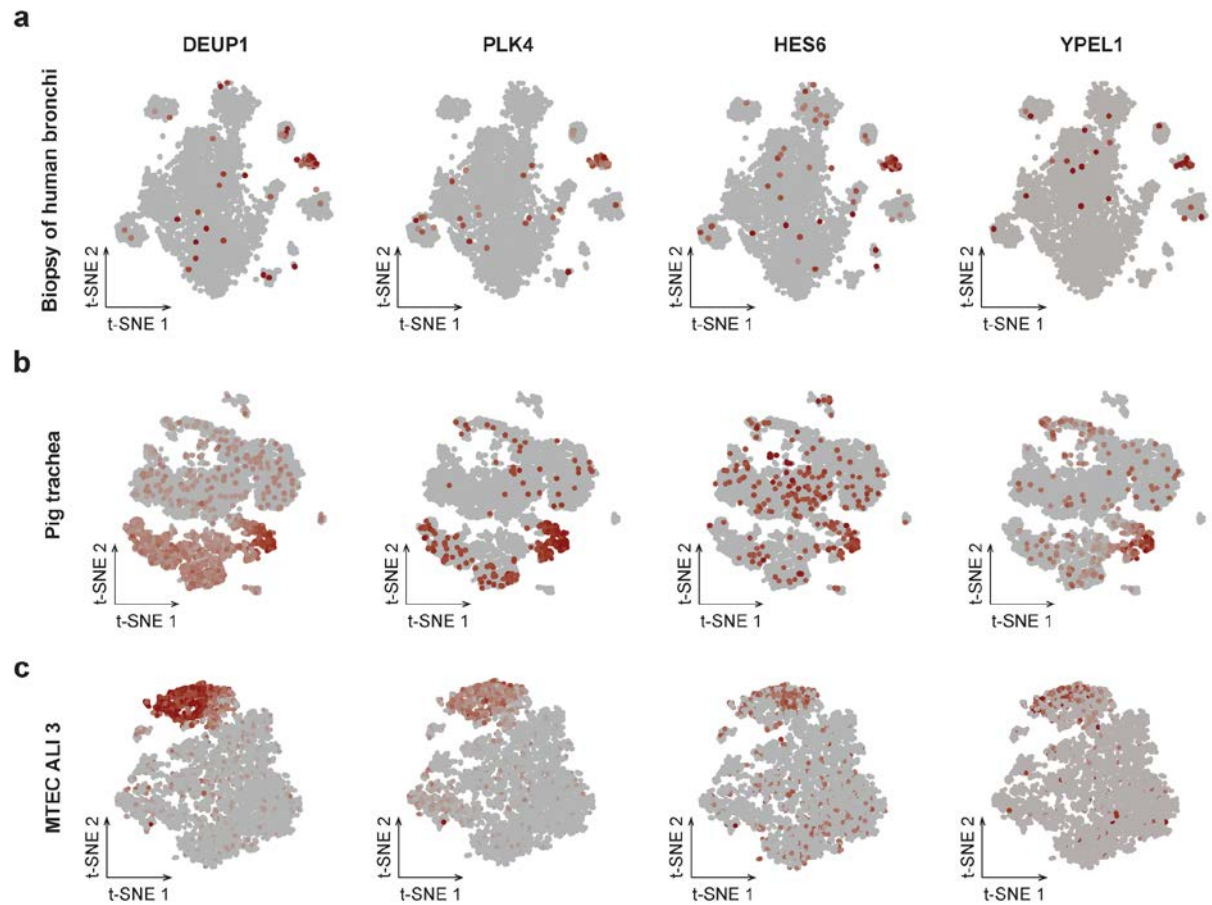
**a** Dendrogram of cell types identified through robust and high-resolution clustering of scRNA-seq data from Pneumacult ALI28 sample. **b** Correlation matrix of cell type identified in Pneumacult sample. **c** t-SNE representation of scRNA-seq data from human nasal brushing (left), nasal turbinate (center), bronchial biopsy (right) colored by cell type described previously (Fig. 1), in grey are rare or non-epithelial cell types. **d** t-SNE representation of scRNA-seq data from pig trachea (left) and mouse culture (right) colored by cell type previously described (Fig. 1).





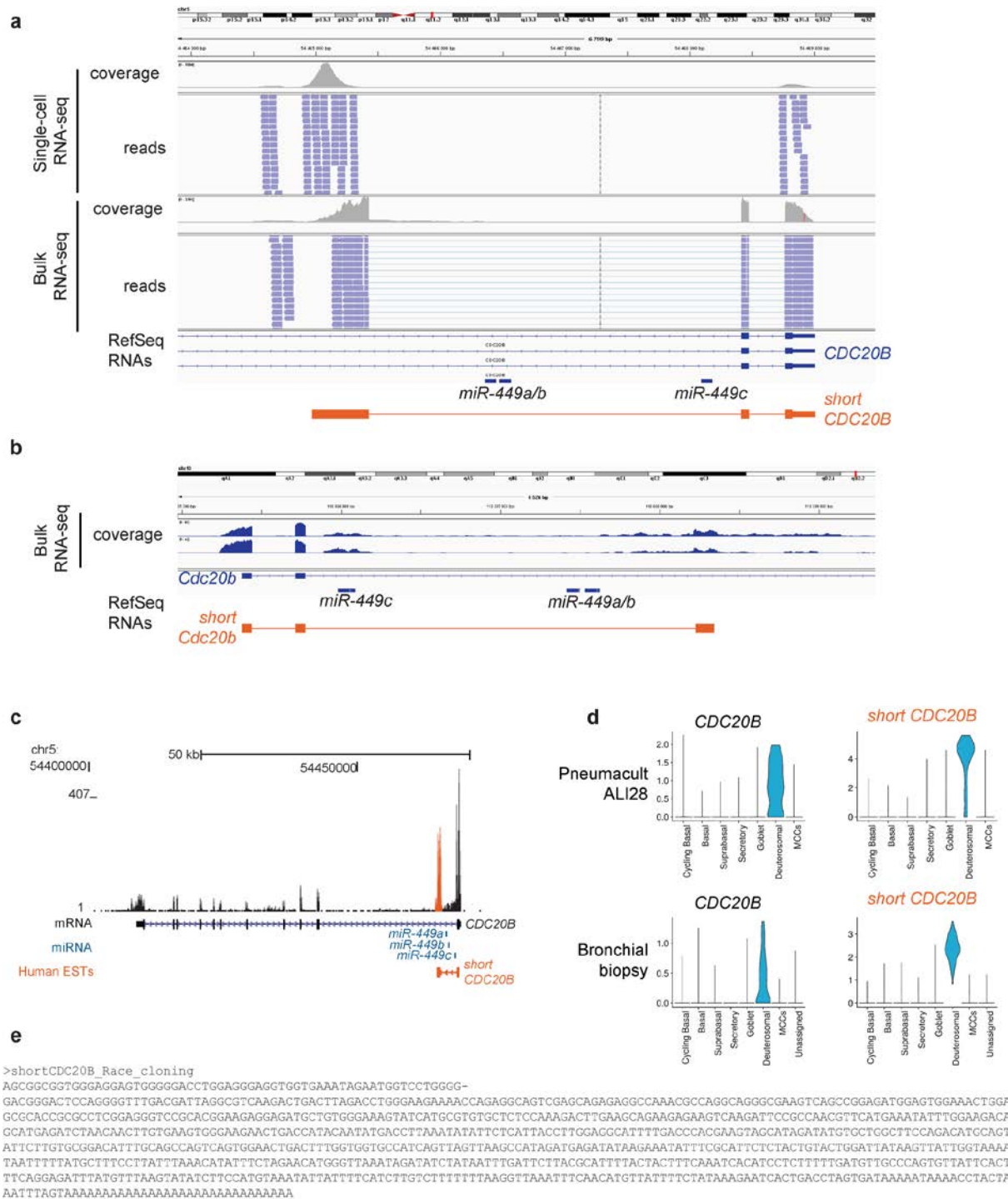
**Supplementary figure 7: Pathway enrichment comparison between basal, suprabasal and secretory cell subtypes**

Identification, with Ingenuity Pathway Analysis (Qiagen) of the enriched canonical pathways, upstream regulators and disease and functions in each of the basal and secretory subpopulations, based on an analysis of 300 differential expressed genes (FDR<0.01 and log2FC>0.6 for BCs and SCs, and log2FC>0.4 for supraBCs). C: cycling basal cells.



### Supplementary figure 8: Robustness of Deuterosomal cells marker genes

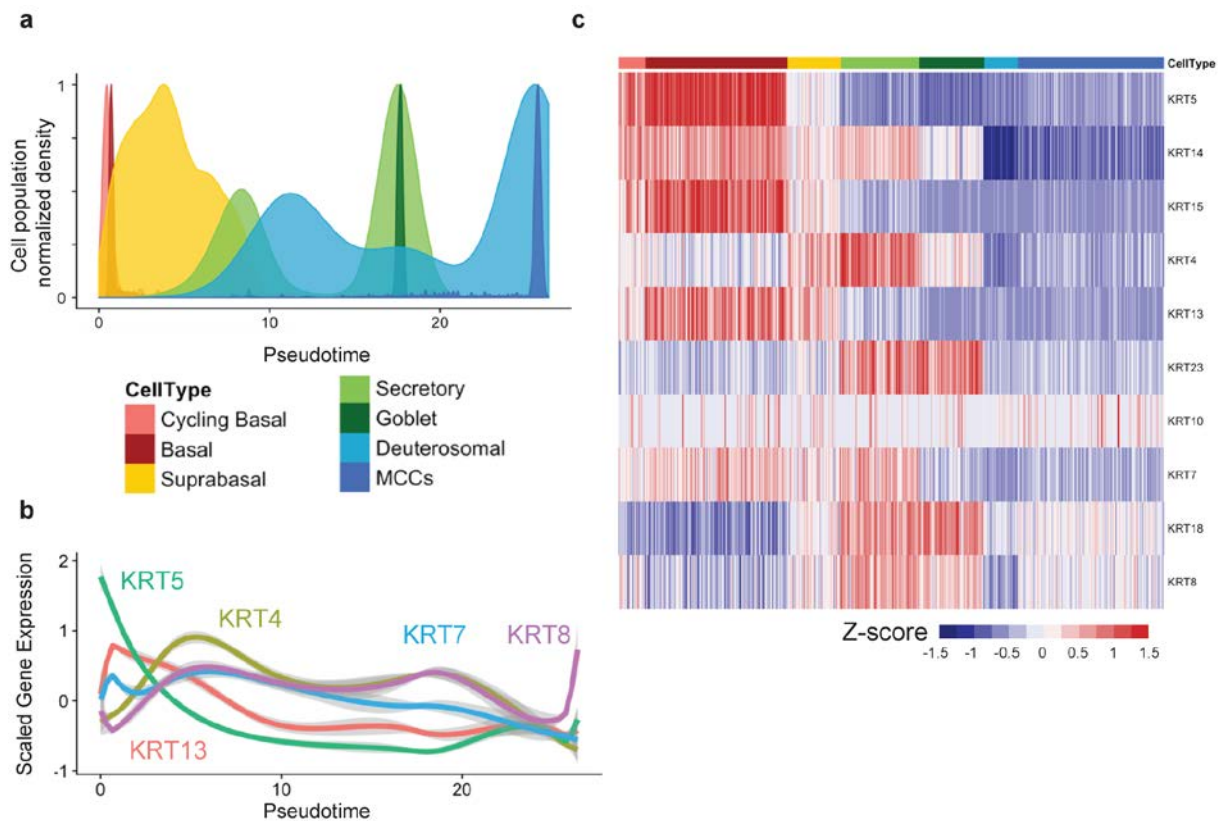
**a, b, c** t-SNE representation of Deuterosomal cell population marker genes expression (lowly to highly expressed, grey to red) in (a) Biopsy of human trachea, (b) Pig trachea, (c) MTECs ALI 3.



**Supplementary figure 9: Identification of short CDC20B, a novel isoform of a deuterosomal cell population marker gene.**

**a** Integrated Genome Viewer (hg19) view of the CDC20B gene with coverage and read alignment from bulk RNA-seq of well-differentiated HAECs and single-cell RNAseq of ALI14 differentiated HAECs. **b** Genome Viewer (mm10) view of the CDC20B gene with coverage from bulk RNA-seq of ALI 7 differentiated MTECs from the public dataset GSE75715. **c** UCSC Genome browser (hg19) view of the CDC20B gene with coverage from bulk RNA-seq of ALI28 differentiated HAECs.

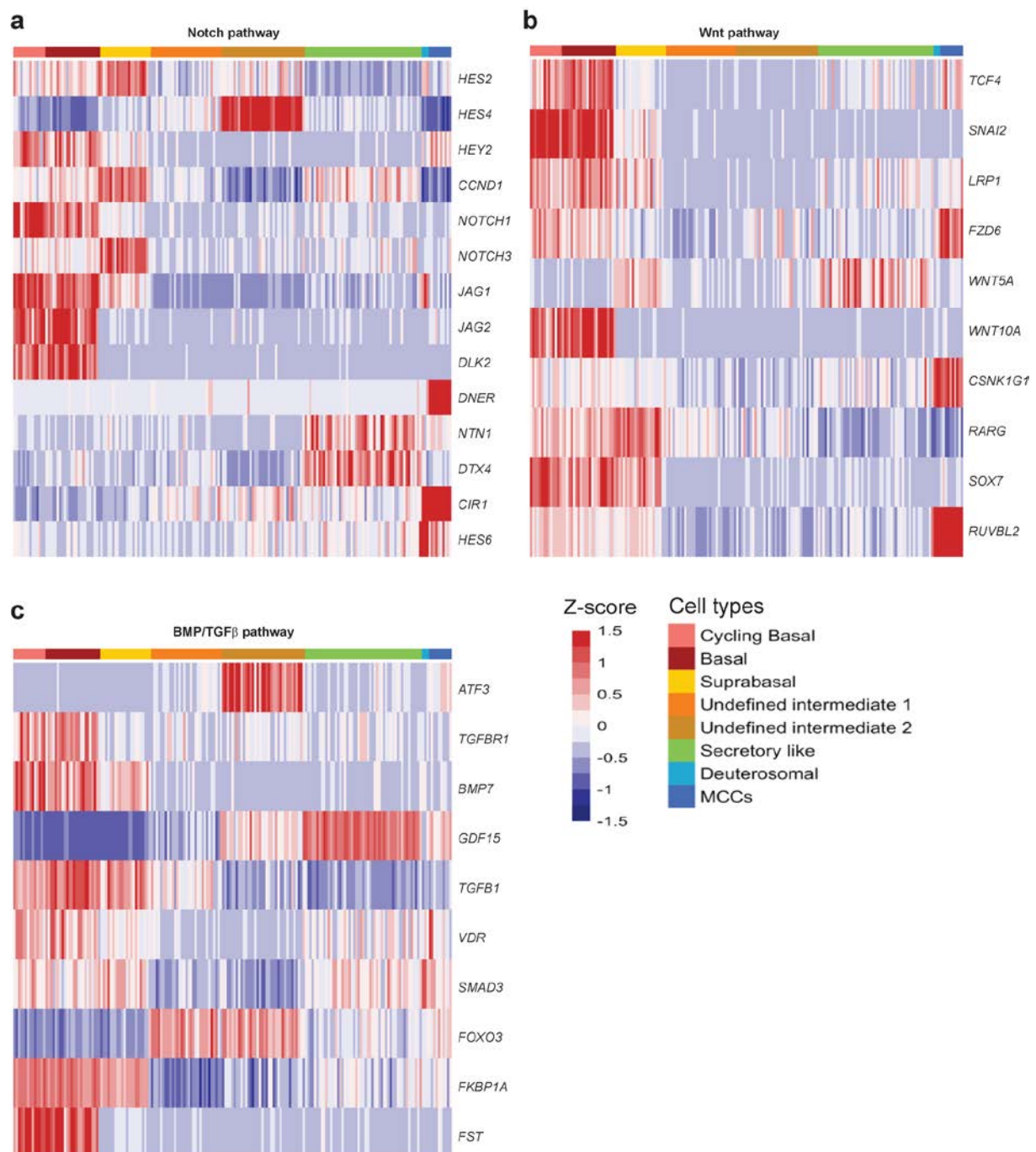
The alternative 3rd exon is shown in orange. **d** Violin plots for CDC20B and short CDC20B abundance in the Pneumacult ALI28 and the bronchial biopsies samples. **e** Alternative human short CDC20B sequence identified with 5' race cloning.



### Supplementary figure 10: Keratin expression in scRNA-seq from pig tracheal cells

**a** Distribution of the 7 main cell types in the pseudotime from scRNA-seq of pig tracheal cells. **b** Plot of normalized gene expression of selected keratins according to pseudotime. **c** Heatmap representing the smoothed temporal expression pattern of regulated keratins. Cells were ordered by cluster appearance.





**Supplementary figure 11: Single-cell expression of signaling pathways components during airway regeneration in BEGM medium**

**a** Heatmap of the genes related to the NOTCH pathway with cells ordered by cluster. **b** Heatmap of the genes related to the WNT pathway with cells ordered by cluster. **c** Heatmap of the genes related to the BMP/TGF $\beta$  pathway with cells ordered by cluster.

## Supplementary tables

**Supplementary table 1: Differential Gene expression of Deuterosomal cluster vs. All clusters**

gene	avg_logFC	pct.1	pct.2	p_val_adj
CDC20Bshort	4.29173241	0.895	0.106	4.10E-87
CCNO	2.47379533	0.86	0.324	7.42E-28
HES6	2.18385143	0.965	0.112	8.70E-97
C5orf49	2.05693783	0.877	0.24	5.15E-36
PSIP1	2.01390765	0.982	0.205	1.19E-61
CCDC34	1.9809765	0.895	0.335	6.78E-32
KIF9	1.94960808	0.947	0.349	4.93E-34
TPPP3	1.92875042	0.825	0.308	4.35E-23
ROPN1L	1.89011241	0.789	0.206	1.05E-30
PSENEN	1.86666232	0.965	0.624	2.30E-23
PCM1	1.84248476	1	0.64	6.52E-32
CCDC74A	1.84041155	0.825	0.211	1.13E-34
TUBB4B	1.81532629	0.965	0.795	1.36E-19
FOXJ1	1.80998137	0.947	0.097	7.16E-97
C20orf85	1.79683265	0.86	0.379	4.73E-19
C1orf192	1.73404906	0.86	0.136	1.11E-54
C11orf88	1.7232126	0.754	0.192	9.00E-30
ZMYND10	1.71516915	0.807	0.136	6.82E-50
RSPH9	1.71414237	0.754	0.12	4.12E-48
LRRIQ1	1.71044455	0.912	0.15	6.63E-58
HSP90AA1	1.70454418	1	0.997	2.63E-21
WDR54	1.67122137	0.702	0.215	1.25E-22
EFCAB1	1.65220289	0.754	0.156	3.48E-36
TUBA1B	1.64767051	1	0.692	2.18E-27
FAM183A	1.64248298	0.807	0.253	4.34E-25
C9orf9	1.6414399	0.772	0.133	2.01E-46
FAM229B	1.63041554	0.877	0.402	8.13E-21
C9orf116	1.62919072	0.93	0.283	2.56E-33
TCTEX1D2	1.62395152	0.93	0.496	5.39E-22



NUCB2	1.61638026	0.982	0.674	4.85E-28
MNS1	1.6004758	0.807	0.103	2.81E-63
MUC12	1.59921724	0.667	0.086	7.39E-51
NUDC	1.59651915	0.965	0.648	2.25E-24
NDUFAF3	1.59393536	0.965	0.712	3.13E-20
EVL	1.57890572	0.667	0.307	7.39E-11
RPL39L	1.56640794	0.772	0.228	1.49E-29
ZDHH1C	1.51682303	0.579	0.108	8.37E-29
CCDC19	1.51575188	0.807	0.102	1.86E-64
RIBC2	1.51195731	0.772	0.079	2.82E-76
CCDC104	1.51138797	0.842	0.459	1.18E-18
CAPSL	1.50462124	0.737	0.2	1.03E-25
CCDC146	1.5016302	0.807	0.264	9.31E-25
BTG3	1.47849706	0.912	0.486	7.95E-24
CALM3	1.47740454	0.912	0.704	3.29E-18
CEP78	1.47525634	0.842	0.125	2.65E-63
RP11-620J15.3	1.46515404	0.789	0.201	6.34E-31
C22orf15	1.44376914	0.702	0.139	1.86E-32
ODF2L	1.44349845	1	0.663	1.77E-25
CCDC170	1.4402005	0.754	0.142	4.32E-37
RAMP1	1.42617804	0.947	0.324	1.54E-29
C1orf194	1.41626622	0.719	0.187	6.75E-24
RABL5	1.41085988	0.895	0.411	1.59E-23
RSPH1	1.40994819	0.86	0.302	6.72E-24
HSPH1	1.39112603	0.86	0.484	9.44E-18
PIFO	1.38292969	0.86	0.198	1.19E-36
RGCC	1.37321897	0.702	0.428	2.17E-08
HIST1H1C	1.36241342	0.842	0.63	2.97E-10
RP11-356K23.1	1.35863787	0.789	0.297	1.14E-17
RUVBL2	1.35693629	0.877	0.361	1.01E-24
BPIFB1	1.34577944	0.965	0.712	9.42E-16

CBY1	1.34356571	0.86	0.436	3.44E-20
IDH2	1.34082721	0.912	0.482	6.90E-22
DYX1C1	1.3387954	0.754	0.131	1.31E-43
CCDC181	1.33688299	0.789	0.124	2.37E-52
CCNA1	1.3354672	0.544	0.175	3.80E-13
FAM81B	1.33244038	0.737	0.098	3.55E-51
CALM2	1.32413545	1	0.983	8.24E-22
NPHP1	1.3177589	0.842	0.136	1.11E-55
IQCG	1.30701982	0.789	0.17	1.34E-36
ENKUR	1.29320702	0.772	0.103	1.47E-54
PTGES3	1.2881322	0.947	0.704	2.56E-22
B9D2	1.28680325	0.754	0.203	3.48E-28
C21orf59	1.28678643	0.93	0.429	1.85E-25
HSPA1A	1.28654824	0.737	0.393	4.02E-12
LRRC26	1.28611747	0.807	0.156	4.97E-44
CDK2AP2	1.27190382	0.86	0.522	1.43E-15
CARHSP1	1.27014247	0.947	0.542	8.33E-21
AKR7A2	1.25970564	0.86	0.549	1.18E-16
MORN2	1.25529358	0.947	0.579	4.71E-21
IFT27	1.25311043	0.807	0.457	3.00E-14
MLF1	1.24180441	0.789	0.17	6.30E-37
C11orf74	1.24151256	0.772	0.27	1.94E-21
TNFAIP8L1	1.23511791	0.807	0.119	4.50E-53
IFT57	1.23511155	0.895	0.435	2.20E-20
PPP1R2	1.22361833	0.807	0.502	3.22E-14
SPA17	1.22298091	0.842	0.255	5.24E-28
REPIN1	1.21730845	0.702	0.207	1.75E-22
TMEM106C	1.21602546	0.912	0.34	5.84E-30
CAPS	1.21510639	0.93	0.596	3.80E-17
CENPF	1.20650472	0.737	0.051	1.56E-100
AGR3	1.2009874	0.684	0.158	2.91E-25
DNALI1	1.19420228	0.789	0.144	7.65E-42
COPRS	1.19044873	0.895	0.428	1.51E-23

DNAAF1	1.18760976	0.684	0.087	1.39E-45
DNAH12	1.18723425	0.684	0.077	2.35E-55
P4HTM	1.17749297	0.807	0.208	7.23E-33
CDS1	1.1651768	0.789	0.304	3.14E-19
CCDC176	1.16312236	0.825	0.107	1.13E-60
CCDC67	1.16206565	0.561	0.026	1.26E-104
DYNLL1	1.16095038	1	0.998	2.61E-18
PPIL6	1.15746657	0.719	0.11	6.89E-43
SPAG6	1.15571312	0.649	0.077	7.90E-51
NEK2	1.14382991	0.667	0.01	1.37E-253
C14orf142	1.12998563	0.737	0.281	1.73E-16
DPCD	1.11695438	0.86	0.418	3.67E-18
MRPL43	1.11527385	0.807	0.743	6.40E-05
IFT43	1.11522388	0.825	0.491	3.92E-14
C11orf70	1.11311121	0.702	0.141	8.27E-33
CKS1B	1.10982697	0.947	0.516	6.97E-25
SMC4	1.10960774	0.807	0.275	3.38E-24
ENDOG	1.10892823	0.737	0.29	2.13E-16
IK	1.10174504	0.877	0.562	3.67E-15
CDC20	1.08621046	0.649	0.028	1.21E-130
SPAG1	1.08478524	0.842	0.218	6.73E-33
TRAF3IP1	1.08278447	0.737	0.231	1.50E-22
EFHC1	1.07961279	0.842	0.222	2.65E-30
EIF2S2	1.07650973	0.965	0.94	4.59E-20
TEKT1	1.07591039	0.649	0.11	4.64E-35
TUBA1A	1.07269853	0.754	0.401	4.14E-10
SPATA33	1.07039339	0.86	0.198	2.55E-40
C7orf57	1.06940792	0.596	0.053	4.15E-62
TXNRD1	1.06515922	0.842	0.368	2.66E-20
PLCE1	1.06316626	0.719	0.155	1.19E-31
LRRC23	1.06274825	0.789	0.203	1.03E-27
HSPA8	1.0625061	0.982	0.897	2.30E-15
MAP1A	1.06179327	0.649	0.074	5.78E-52

MUC1	1.0600729	0.965	0.66	1.18E-15
HSPBP1	1.05799385	0.789	0.255	1.47E-24
SPEF1	1.0573456	0.754	0.086	5.38E-62
CNTRL	1.05449509	0.702	0.151	8.76E-31
TMEM14B	1.05317828	0.947	0.846	2.30E-20
ANKRD36C	1.05092894	0.754	0.299	4.53E-14
YWHAH	1.05078326	0.842	0.468	1.17E-14
H2AFZ	1.05056193	1	0.977	8.47E-26
FANK1	1.04532129	0.702	0.143	3.13E-31
KLHDC9	1.0422098	0.702	0.141	7.43E-33
LRRC46	1.04043533	0.579	0.11	2.40E-26
RANBP1	1.04028579	0.965	0.818	7.39E-22
C9orf135	1.03949632	0.684	0.143	1.23E-27
SNTN	1.03863809	0.596	0.174	1.22E-14
CCDC171	1.0350885	0.684	0.071	1.95E-64
HIST1H2BJ	1.0304779	0.702	0.071	1.98E-70
HSPE1	1.02651432	0.982	0.973	1.98E-17
CDC20B	1.02530067	0.702	0.006	0
LINC01171	1.02417279	0.719	0.122	9.43E-37
H2AFV	1.0211016	0.982	0.775	3.50E-23
CHORDC1	1.01679794	0.772	0.274	1.39E-21
ARMC3	1.01478908	0.719	0.073	3.25E-65
SPAG16	1.01355423	0.825	0.453	3.56E-14
DPY30	1.0099647	0.895	0.793	2.62E-11
MRPL18	1.00932664	0.947	0.776	5.95E-13
CCP110	1.00845073	0.789	0.141	3.63E-43
CEP41	1.00636513	0.684	0.149	7.89E-31
CETN2	1.00537082	0.842	0.553	8.17E-13
ZNF487	1.00212949	0.684	0.101	4.78E-43
DNAJA1	1.00011384	0.93	0.663	3.22E-15

**Supplementary table 2: Primers used in the Biomark qRT-PCR**

Gene	Sequence of forward primer	Sequence of reverse primer
AKAP14	AGATTGTGGAAGAGGAGCGAAACCC	TACCATGCTGGAGTGGTCGGC
CCDC40	CATCCACGGAGTCTTAGGC	TGGATCCTGTCAATCTGCC
CDC20B	CGGCTGAGAAATATGCTTGG	ATAAACACGCCCCAGTCTTG
DNAI1	AACGACGGCTGTCCCTAAAG	AGCCTACAAAACGCTCCCTC
FOXJ1	TGGATCACGGACAACCTTCTG	GAGGCACTTTGATGAAGCAC
IDAS	TTTCAGAGACACGGTGGATG	TGGTGATATGTCGCAAGGAG
KRT5	AGGAATGCAGACTCAGTGGAG	CAGAGGAAACACTGCTTGTGAC
MUC16	GTCAAGCCAGGCAGCACAAGG	GGGATGTGCCTGCTGGACTGC
MUC5AC	CCAAATACGCCAACAAGACC	ATTCCATGGGTGTCAGCTTG
MUC5B	TGCAACCGTCCCGGCTTCG	GTTGTGTTGCACACGCACACCG
MYB	ATGATGAAGACCCTGAGAAGGAAA	AACAGGTGCACTGTCTCCATGA
NEK10	GCAAGAAATCACCATCAGGGAC	GGCTGGAAGCTGTTGTTTGC
PLK4	TCCAACACAGGCACCAATC	GAGATGTCTGTTCCAGAAGCTG
RFX2	GCGACCACATCCTCTACCAG	AGTTACGGATGGCCTGTGTC
RFX3	AGCCAACATCATCAACAGTTTTT	TGCAGTGACTTGATATCTCAAA
ROPN1L	CTCGCATCCCCTTCAAGACG	TCTTCCTGGCGTCTATATTTTCC
SCGB1A1	AGAGACGGGCCAGAGCATCCC	GGCAGATCTCTGCAGAAGCGGAGC
SPRR3	TCTGCACAGCAGGTCCAGCATC	AGGCTGGCTGGGTTGTTTCACC
TFF3	GGAGTACGTGGGCCTGTCTGC	AGCCCCGGTTGTTGCACTCC