

University of Groningen

Novel views on endotyping asthma, its remission, and COPD

Carpaij, Orestes

DOI:
[10.33612/diss.136744640](https://doi.org/10.33612/diss.136744640)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Carpaij, O. (2020). *Novel views on endotyping asthma, its remission, and COPD*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.136744640>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Chapter 8

A cellular census of human lungs identifies novel cell states in health and in asthma



Felipe A. Vieira Braga [#], Gozde Kar [#], Marijn Berg [#], Orestes A. Carpaij [#], Krzysztof Polanski, Lukas M. Simon, Sharon Brouwer, Tomás Gomes, Laura Hesse, Jian Jiang, Eirini S. Fasouli, Mirjana Efremova, Roser Vento-Tormo, Carlos Talavera-López, Marnix R. Jonker, Karen Affleck, Subarna Palit, Paulina M. Strzelecka, Helen V. Firth, Krishnaa T. Mahbubani, Ana Cvejic, Kerstin B. Meyer, Kouros Saeb-Parsy, Marjan Luinge, Corry-Anke Brandsma, Wim Timens, Ilias Angelidis, Maximilian Strunz, Gerard H. Koppelman, Antoon J. van Oosterhout, Herbert B. Schiller, Fabian J. Theis, Maarten van den Berge, Martijn C. Nawijn ^{*}, Sarah A. Teichmann ^{*}

[#]: shared first author

^{*}: shared last author

Nat Med. 2019 Jul;25(7):1153-1163

Abstract

Human lungs enable efficient gas exchange and form an interface with the environment, which depends on mucosal immunity for protection against infectious agents. Tightly controlled interactions between structural and immune cells are required to maintain lung homeostasis. Here, we use single-cell transcriptomics to chart the cellular landscape of upper and lower airways and lung parenchyma in healthy lungs, and lower airways in asthmatic lungs. We report location-dependent airway epithelial cell states and a novel subset of tissue-resident memory T cells. In the lower airways of patients with asthma, mucous cell hyperplasia is shown to stem from a novel mucous ciliated cell state, as well as goblet cell hyperplasia. We report the presence of pathogenic effector type 2 helper T cells (TH2) in asthmatic lungs and find evidence for type 2 cytokines in maintaining the altered epithelial cell states. Unbiased analysis of cell-cell interactions identifies a shift from airway structural cell communication in healthy lungs to a TH2-dominated interactome in asthmatic lungs.

Main

The lung plays a critical role in both gas exchange and mucosal immunity, and its anatomy serves these functions through: (1) the airways that lead air to the respiratory unit, provide mucociliary clearance and form a barrier against inhaled particles and pathogens; and (2) the alveoli, distal saccular structures where gas exchange occurs. We set out to profile lung-resident structural and inflammatory cells and their interactions by analyzing healthy human respiratory tissue from four sources: nasal brushes, bronchial biopsies and brushes from living donors, tissue samples from lung resections and transplant donor lungs. Our single-cell analysis identifies differences in the proportions and transcriptional phenotype of structural and inflammatory cells between upper and lower airways and lung parenchyma. We identify a novel tissue-resident CD4 T cell subset that harbors features of both circulating memory cells and of tissue-resident memory (TRM) cells. We demonstrate that many disease-associated genes have highly cell type-specific expression patterns.

In addition, we evaluate the altered cellular landscape of the airway wall in asthmatic lungs. We identify a novel epithelial cell state highly enriched in asthmatic lungs. These mucous ciliated cells represent a transitioning state of ciliated cells and contribute to mucous cell hyperplasia in this chronic disease. Other asthma-associated changes include increased numbers of goblet cells, intraepithelial mast cells, and pathogenic effector type 2 helper T cells (TH2) in airway wall tissue. Analysis of intercellular communications in healthy and asthmatic airway walls reveals a remarkable loss of structural cell communication and a concomitant increase in TH2 cell interactions. We generate novel insights into epithelial cell changes and altered communication patterns between immune and structural cells of the airways that underlie asthmatic airway inflammation.

Results

Human lung cell census identifies macro-anatomical patterns of epithelial cell states across the human respiratory tree

The cellular landscape along the 23 generations of the airways in human lung is expected to differ both in terms of relative frequencies of cell types and their molecular phenotype [1]. We used 10x Genomics Chromium droplet single-cell RNA sequencing (scRNA-seq) to profile a total of 36,931 single cells from upper and lower airways and lung parenchyma (Fig. 1a,b and Supplementary Table 1). We profiled nasal brushes, bronchial brushes, and airway wall biopsies (third to sixth generation) from healthy volunteers. For parenchyma (small respiratory airways and alveoli), we obtained lung tissue from deceased transplant donors.

Our analysis reveals a diversity of epithelial, endothelial, stromal, and immune cells, with approximately 21 coarse-grained cell types in total (Figs. 1 and 2 and Extended Data Fig. 1), that can be explored in a user-friendly web portal (www.lungcellatlas.org). We further confirmed our observations by comparing our dataset with parenchymal lung tissue from resection material analyzed on a bespoke Drop-seq-like microfluidics platform [2] (Extended Data Fig. 2). We observed extensive overlap in cell-type identities (Extended Data Fig. 2). In our analysis below, we first concentrate on epithelial cells (Fig. 1) and then focus on the stromal and immune compartments (Fig. 2).

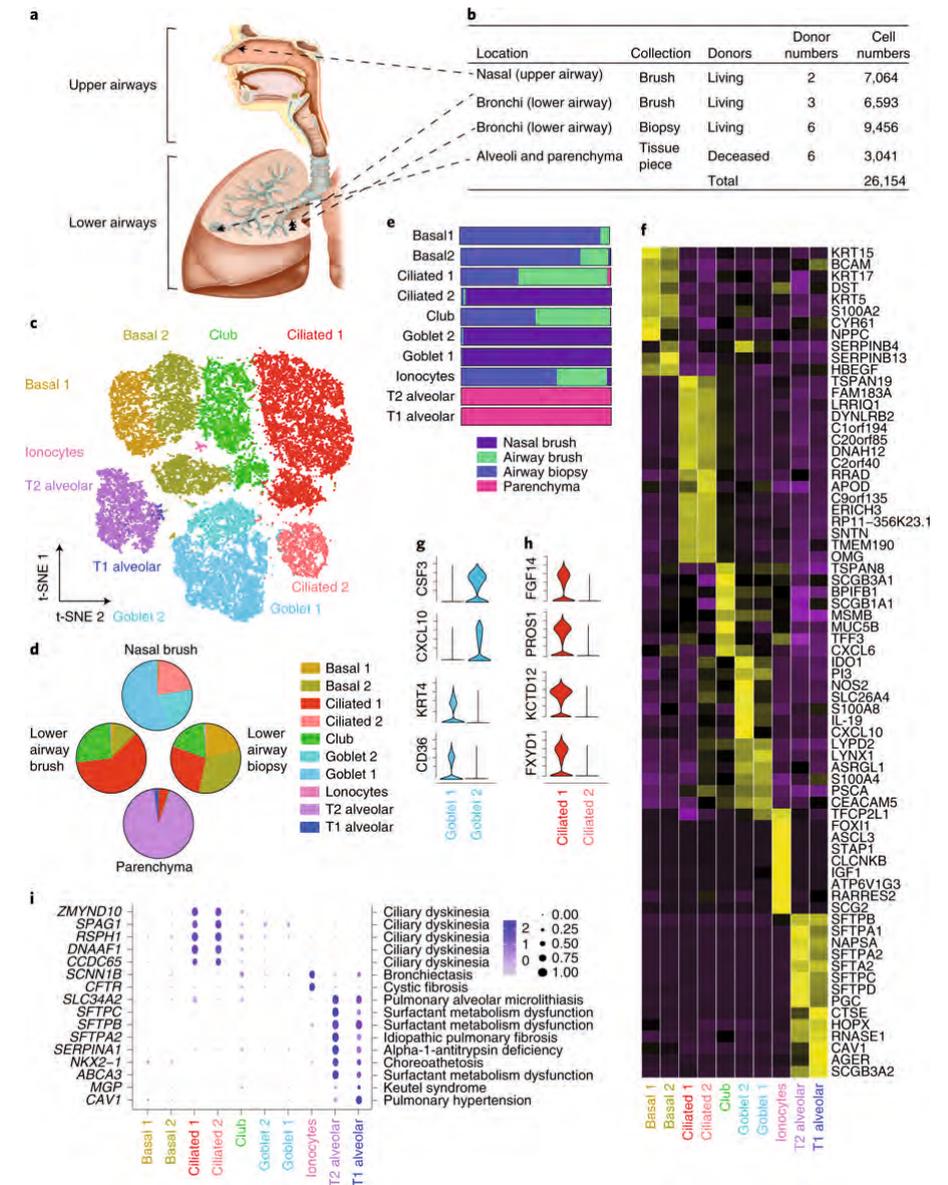


Figure 1: A human lung cell census identifies zonation of novel epithelial cell states across macro-anatomical location. **a:** schematic depicting anatomical regions analyzed in this study, **b:** table with the details of anatomical region, tissue source, donors, and cell numbers present in this figure, **c:** t-SNE displaying the major epithelial clusters present in the full extent of the human respiratory tree. T1, type 1; T2, type 2, **d:** Pie charts depicting the cellular composition by anatomical region, **e:** Horizontal slice bar depicting the anatomical distribution of each cell type identified, **f:** Heat map depicting the average expression levels per cluster of the top differentially expressed markers in each cluster, **g:** Violin plots of selected markers identified by differential expression analysis comparing the two goblet subsets to each other, **h:** Violin plots of selected markers identified by differential expression analysis of ciliated 1 versus ciliated 2 clusters, **i:** Dot plot depicting gene expression levels and percentage of cells expressing genes associated with specific lung phenotypes according to the OMIM database. All the differential expression analyses in f–i were performed using the non-parametric two-sided Wilcoxon rank sum test in Seurat. All panels depict the number of cells and individuals described in b.

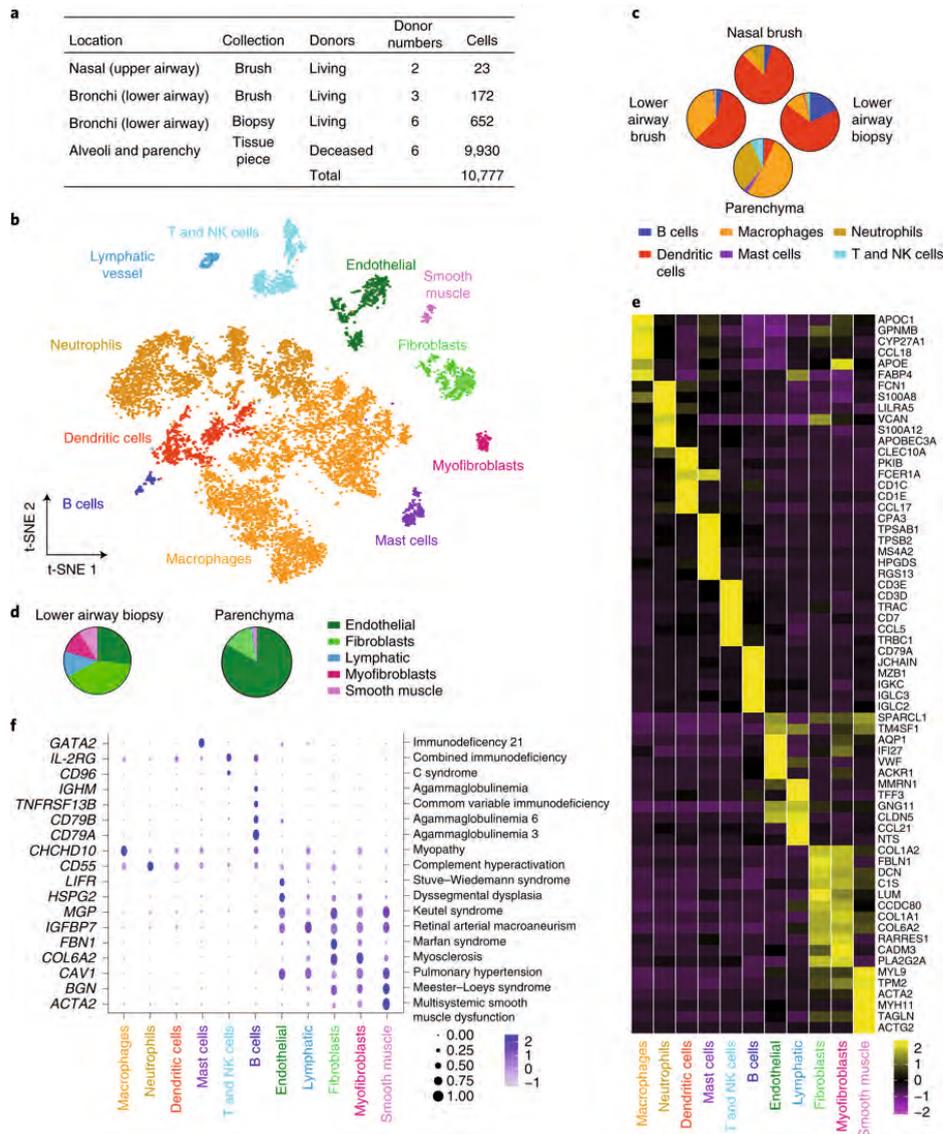


Figure 2: A cellular and molecular map of the stromal and immune components of across the upper and lower human respiratory airways. **a:** Table with details of anatomical region, tissue source, donors, and cell numbers present in this figure, **b:** t-SNE displaying the major immune and mesenchymal clusters present in the full extent of the human respiratory tree. NK, natural killer, **c:** Pie charts depicting the cellular composition of immune cells by anatomical region, **d:** Pie charts depicting the cellular composition of stromal cells in lower airway biopsies and parenchyma tissue, **e:** Heat map depicting the average expression levels per cluster of the top differentially expressed markers in each cluster, **f:** Dot plot depicting gene expression levels and percentage of cells expressing genes associated with lung phenotypes according to the OMIM database. All the differential expression analyses in **e** and **f** were performed using the non-parametric two-sided Wilcoxon rank sum test in Seurat. All panels depict the number of cells and individuals described in **a**.

In the epithelial lineage, we identified at least ten cell types across the upper and lower airways and lung parenchyma (Fig. 1c and Extended Data Fig. 1). We detected multiple basal, club, ciliated, and goblet cell states, as well as type 1 and type 2 alveolar cells, and the recently described ionocyte [3,4] (Fig. 1d and Extended Data Fig. 3). We did not identify specific clusters of tuft cells or neuroendocrine cells. Supervised analysis using neuroendocrine (CHGA, ASCL1, INSM1, HOXB5) and tuft (DCLK1, ASCL2)[3] cell marker genes identified a small number of cells with neuroendocrine-like features present only in lower airways (Extended Data Fig. 4). Tuft cell marker genes did not identify a unique cell population.

We identified two discrete cell states in each of basal, goblet, and ciliated epithelial cells. Basal cells were present in upper and lower airways, although at relatively low frequency in upper airways (Fig. 1e). The two basal cell states corresponded to differentiation stages, with the less-mature basal 1 cell state expressing higher levels of TP63 and NPPC compared with more-mature basal 2 cells (Fig. 1f and Extended Data Fig. 1), which were more abundant in bronchial brushes, suggesting a more apical localization (Fig. 1d,e). Goblet 1 and 2 cells were both characterized by high expression of CEACAM5, S100A4 and MUC5AC, and lack of MUC5B (Fig. 1f and Extended Data Figs. 1 and 4). Goblet 1 cells specifically express KRT4 and CD36 (Fig. 1g and Extended Data Fig. 4). Genes involved in recruitment of neutrophils, monocytes, dendritic cells, and T cells, such as IDO1, NOS2, IL19, CSF3 (granulocyte-colony stimulating factor), and CXCL10 are expressed at high levels in goblet 2 cells (Fig. 1f,g and Extended Data Fig. 4). Both goblet cell states are present in upper airway epithelium. In contrast, only the goblet 2 cell state was present in lower airways, albeit at low abundance (Fig. 1e).

Ciliated cells were also zoned in terms of their presence across macro-anatomical locations, with a discrete ciliated cell state more abundant in upper airways (ciliated 2) compared with lower airways and lung parenchyma. Nasal epithelial ciliated 2 cells express proinflammatory genes, such as CCL20, and higher levels of metabolic (ATP12A and COX7A1) and vesicle transport (AP2B1 and SYT5 [6]) genes compared with the ciliated 1 cells (Extended Data Fig. 4). In contrast, the ciliated 1 cells from lower airways specifically expressed genes involved in cytoprotection (PROS1 [7]) and fluid reabsorption (FXR1 [8]) (Fig. 1h and Extended Data Fig. 4). We detected a location-

specific transcriptional signature specific for the upper airways in both ciliated and goblet cells (Extended Data Fig. 4b).

Next, we assessed the contribution of specific epithelial cell types to Mendelian disease. Cell type-specific expression patterns of genes associated with Mendelian disorders (based on the Online Mendelian Inheritance in Man (OMIM) database) confirmed ionocytes as particularly high expressors of the CFTR gene, mutated in cystic fibrosis (Fig. 1i and Supplementary Table 2). These cells also express SCNN1B, mutations of which can cause bronchiectasis, another feature of cystic fibrosis, suggesting a pathological role for ionocytes in both bronchiectasis and cystic fibrosis. In addition, expression of SERPINA1 (Fig. 1i) was found to be enriched in type 2 alveolar epithelial cells, underscoring their role in alpha-1-antitrypsin deficiency [9].

Differential anatomical distribution of the stromal and immune components in the human respiratory tree

Next, we analyzed the single-cell transcriptomes of immune and stromal cells from the upper airways, lower airways, and the lung parenchyma (Fig. 2a). We identified immune clusters of myeloid (macrophages, neutrophils, dendritic cells, and mast cells) and lymphoid cells (T and natural killer cells, B cells; Fig. 2b and Extended Data Fig. 5). Immune and stromal cell numbers and composition varied greatly across different anatomical regions (Fig. 2a,c). Nasal brushes contained only a small number of immune cells, with the large majority being dendritic cells. In the lower airways, the fraction of inflammatory cells was much larger and relatively enriched for macrophages (Fig. 2c and Extended Data Fig. 5), which was directly confirmed by cell composition comparison of upper versus lower airway brushes obtained from the same donor (Extended Data Fig. 5e). The stromal component of the parenchyma region was dominated by vascular endothelial cells with a small number of fibroblasts, while the upper airways had a richer stromal compartment, with fibroblasts, myofibroblasts, smooth muscle, vascular, and lymphatic endothelial cells (Fig. 2d).

Macrophage transcriptional phenotypes showed large donor variation (Extended Data Fig. 5), but they all shared high expression of MARCO, CCL18, and genes involved in apolipoprotein metabolism (APOC1 and APOE) (Fig. 2e and Extended Data Fig. 5). Lung neutrophils express high levels of the granulocyte markers S100A8, S100A12 [10], and

LILRA5, a receptor poorly characterized in the lungs that has been shown to have a proinflammatory function in synovial fluid macrophages [11] (Fig. 2e and Extended Data Fig. 5). Dendritic cells were mostly myeloid, with high expression of CD1E, CD1C, and CLEC10A (Fig. 2e and Extended Data Fig. 5) and of FCER1A (IgE receptor) and CCL17, which have a key role in inflammatory conditions such as asthma [12].

In the scRNA-seq datasets, we could not distinguish T cells and natural killer cells from each other (Fig. 2b). The B cells in our dataset were mostly plasma cells, expressing high levels of JCHAIN. IgM⁺ (IGHM) B cells were enriched in the bronchial brushes and in the lung parenchyma, while IgG3⁺ (IGHG3) B cells were enriched in airway biopsy samples and were virtually absent from the bronchial brushes, suggesting a micro-anatomical segregation of B cell subsets (Extended Data Fig. 5f). The immune and stromal lung components also displayed cell type-specific expression patterns of genes associated with Mendelian disorders. However, in contrast to epithelial cell associated syndromes, syndromes associated with immune and stromal cells were largely systemic in nature (with lung involvement [13]) (Fig. 2f).

Molecular features of mucous cell metaplasia in asthma

Asthma is a complex and heterogeneous disease [14], where epithelial [15,16], stromal [17], and immune cells [18,19] contribute to a spectrum of clinical phenotypes. We analyzed bronchial biopsies from six volunteers with chronic, childhood-onset asthma (Fig. 3a and Supplementary Table 3) and compared them with bronchoscopy samples obtained from healthy sex- and age-matched volunteers (Supplementary Table 3). The Global Initiative for Asthma (GINA) score at time of recruitment indicated patients had either mild or moderate asthma (Supplementary Table 3). Withdrawal of inhaled corticosteroids was a minimum of six weeks before sample collection for all patients. Most patients had controlled asthma on inhaled corticosteroid withdrawal (asthma control questionnaire score < 1.5). The combined airway wall dataset reveals a cellular landscape dominated by epithelial cells, with minor contributions from endothelial, mesenchymal, and immune cells (Extended Data Fig. 6a,b and Supplementary Tables 4 and 5).

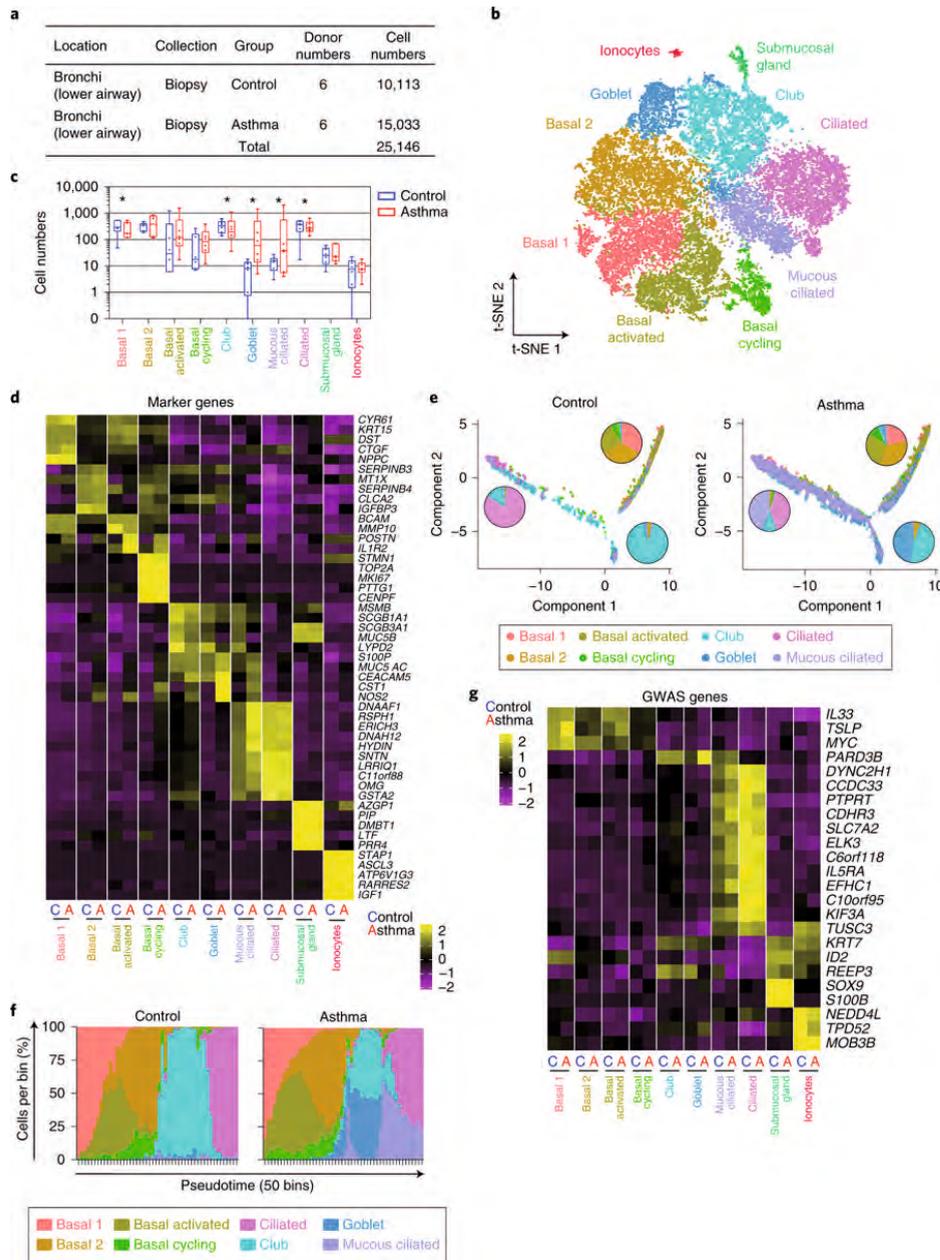


Figure 3: distinct programs of epithelial cell differentiation in asthmatic versus healthy airways, **a:** Table with overview and cell numbers for healthy control volunteers and volunteers with asthma analyzed in this figure, **b:** t-SNE displaying all epithelial cells analyzed colored by their specific cluster assignment, **c:** Box and whisker plots depicting cell numbers of healthy control patients and patients with asthma in each cluster, **d:** Heat map displaying the top five differentially expressed genes per cluster, **e:** Pseudotime developmental trajectory analysis from Monocle2 depicting how each of the basal, secretory, and ciliated subsets relate to each other, **f:** Binned pseudotime analysis displaying how each subset is

ordered in a one-dimensional continuous space, **g:** Heat map displaying the expression of asthma genes from GWAS. Only genes present in our list of differentially expressed genes are depicted for each cell cluster. Significance analysis in **c** calculated using Fisher's exact test corrected for multiple comparison using the Bonferroni method. Significance was calculated using all the clusters present in Figs. 3 and 4, which were derived from the same set of samples. * $P < 0.001$. A full list of P values is given in **Supplementary Table 5**. In box and whisker plots in **c**, all points are shown, and the box represents the second and third quartiles and median. All the differential expression analyses in **d** and **g** were performed using the non-parametric two-sided Wilcoxon rank sum test in Seurat. **Panels b–d** and **g** depict the number of cells and individuals described in **a**; **e** and **f** depict the number of individuals described in **a**.

Clustering of the EPCAM⁺ cells identified ten sub-clusters representing the six epithelial cell types observed in healthy airway wall (Fig. 1c), as well as four additional cell states: mucous ciliated cells, activated basal cells, cycling cells, and serous cells from the submucosal glands (Fig. 3b).

Activated basal cells closely resemble basal 1 cells, but also express proinflammatory genes such as POSTN (Fig. 3d). Cycling basal cells are characterized by expression of canonical marker genes of proliferating cells (MKI67 and TOP2A) (Fig. 3d), and this is the only cluster of airway epithelial cells expressing the squamous cell marker KRT13 (Extended Data Fig. 6).

We observe mucous cell hyperplasia in asthma, with a marked increase in goblet cell numbers (Fig. 3c), which are very rare in healthy airway wall biopsies (Fig. 1e). Moreover, the goblet cell transcriptional phenotype is altered in asthma, with strong up regulation of proinflammatory and remodeling genes NOS2, CEACAM5, and CST1 (Fig. 3d). In addition, we identified a strong increase in mucous ciliated cells, a novel cell state highly similar to ciliated cells, but co-expressing a number of mucous genes also observed in goblet cells, including MUC5AC and CEACAM5 and the ciliated genes FOXJ1 and PIFO (Fig. 3d and Extended Data Fig. 6).

To further dissect the inferred differentiation trajectories of epithelial cells in healthy and asthmatic airway walls, we performed pseudotime analysis [20]. This reveals a trajectory starting with basal cell subsets, bifurcating into either a secretory lineage (mainly club cells) and a ciliated lineage in healthy airway wall (Fig. 3e). In asthma, the secretory lineage is a mix of club and goblet cells, while the mucous ciliated cell state is mapped to the ciliated differentiation trajectory (Fig. 3e,f).

Next, we further analyzed the transcriptional profiles of the mucous ciliated and goblet cells. As both Notch and interleukin-4 (IL-4)/IL-13 signaling by TH2 cells contribute to mucous cell differentiation [21], we analyzed expression of both Notch [22,23] and IL-4/IL-13 target genes [24]. Expression of IL-4/IL-13-induced genes [24] is highest in activated basal cells, goblet cells, and mucous ciliated cells (Extended Data Fig. 7a) and prominent in asthma in club, goblet, and mucous ciliated cells (Extended Data Fig. 7b). In club cells, expression of Notch target genes [22,23] does not differ between asthma and healthy-derived cells. In contrast, in goblet cells, the Notch target gene signature is retained only in cells from healthy airway wall, and is lost in asthma. Interestingly, mucous ciliated cells also lack expression of Notch target genes in asthma (Extended Data Fig. 7). Hence, we postulate that mucous ciliated cells represent a transition cell state in the ciliated lineage—induced by IL-4/IL-13 signaling—leading to a mucous cell phenotype that contributes to mucous cell metaplasia in asthma [21]. Similar to goblet cells, mucous ciliated cells express key asthma genes such as *CST1* [25] and *POSTN* (Fig. 3d), indicating that these cells can contribute to airway inflammation and remodeling.

Analysis of asthma genome-wide association study (GWAS) gene expression in our epithelial scRNA-seq data revealed a broad contribution of airway epithelial cell types to asthma susceptibility (Fig. 3g), with high expression of asthma GWAS genes in ciliated and mucous ciliated cells. This includes genes involved in ciliary function (*DYNC2H1* and *KIF3A*), cell adhesion (*ELK3*, *CDHR3* and *PTPRT*), and IL-5-induced mucus metaplasia (*IL5RA*) [26], further suggesting a direct link between mucous ciliated cells and TH2 cells.

Remodeling of the stromal and immune compartments in asthmatic airways

Asthma is associated with chronic inflammation and remodeling of the airway wall [27]. Analysis of the immune and stromal cell populations (Fig. 4a) in the bronchial biopsies by reveals the presence of B and T cells, neutrophils, macrophages, dendritic cells, mast cells, fibroblasts, smooth muscle cells, and endothelial cells (Fig. 4b and Extended Data Fig. 8). We did not detect innate lymphoid, basophil, or eosinophil clusters (Extended Data Fig. 8). Analysis of bulk transcriptome data of matched airway wall biopsies before and after tissue dissociation identified very low expression levels of the eosinophil marker gene *CLC*, indicating that these cells are rare in the samples we analyzed (Extended Data Fig. 9).

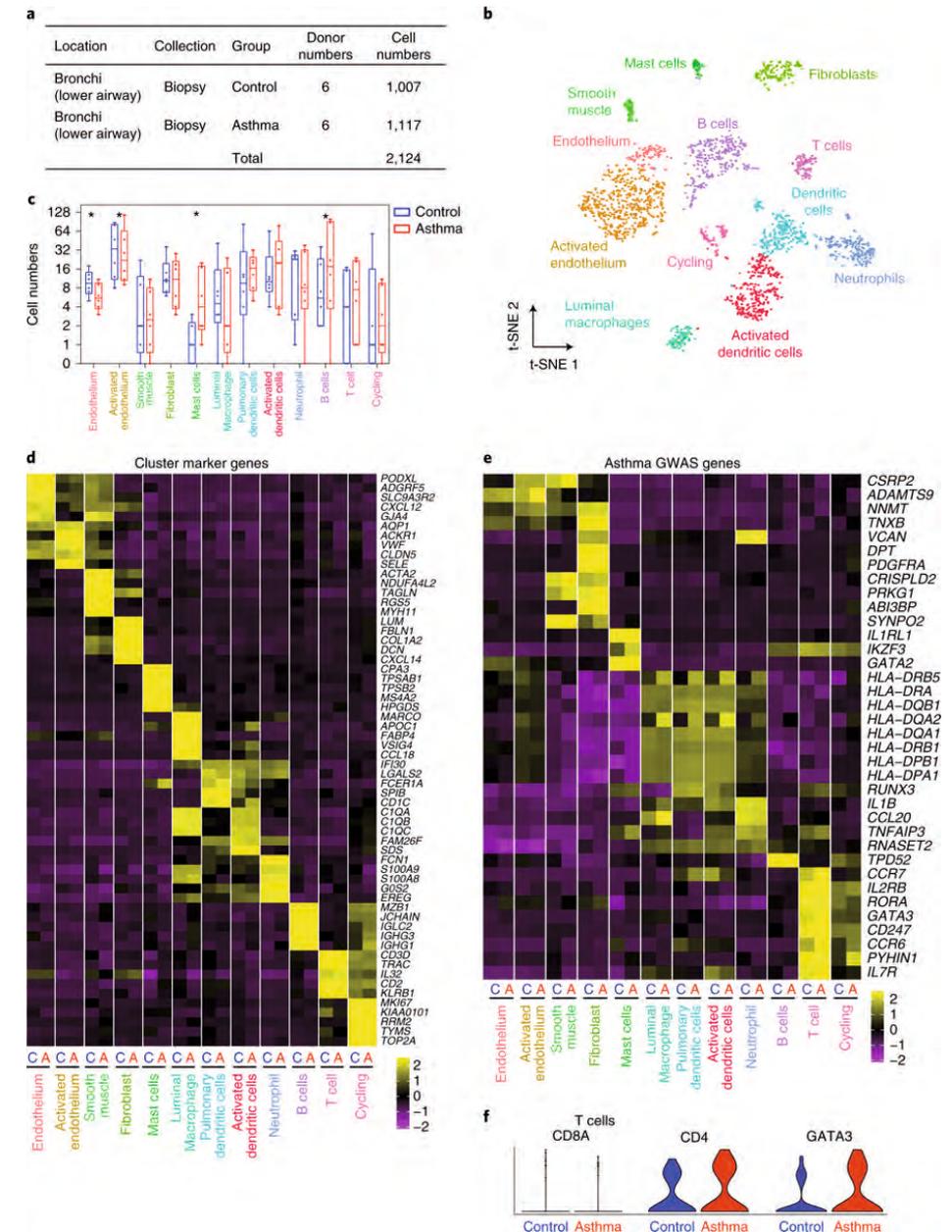


Figure 4: remodeling of the stromal and immune compartments in asthmatic airways, **a:** Table with the number of donors and cells per volunteer group included in this figure, **b:** t-SNE depicting the immune and stromal cell types identified in the human airway combined dataset of healthy controls and patients with asthma, **c:** Box and whisker plots depicting the cell numbers of healthy and asthmatic cells in each cluster, **d:** Heat map displaying gene expression levels of the top five differentially expressed genes per cluster, **e:** Heat map displaying asthma GWAS gene expression per cluster. Only genes present in the top 50 (per cluster) of our list of differentially expressed genes are shown, **f:** Violin plots of selected T cell markers in patients with asthma. Significance was calculated using all the clusters present in **Figs. 3 and 4**, which were

derived from the same set of samples. * $P < 0.001$. A full list of P values is given in **Supplementary Table 5**. All the differential expression analyses in d and f were performed using the non-parametric two-sided Wilcoxon rank sum test in Seurat. In the box and whisker plot in c, all points are shown, and the box represents the second and third quartiles and median. Panels b–f depict the number of cells and individuals described in a.

Mast cell numbers were increased in asthma (Fig. 4c). Mast cells in asthmatic airways lack chymase 1 expression (CMA1) and express high levels of tryptase genes (TPSB2, TPSAB1) and prostaglandin synthetase genes PTGS2 and HPGDS (Fig. 4d and Extended Data Fig. 8d). PTGS2 (cyclooxygenase-2), also known as inflammatory cyclooxygenase, converts the precursor arachidonic acid to prostaglandin endoperoxide H₂ (PGH₂). HPGDS (hematopoietic prostaglandin D synthase) catalyzes the conversion of PGH₂ to prostaglandin D₂ (PGD₂). PGD₂ activates TH₂ cells [28], innate lymphoid cells 2 [29], basophils, and neutrophils [28] and plays a key role in asthma pathology. Expression of all PGD₂ biosynthesis enzymes is a unique feature of mast cells (Extended Data Fig. 8d), suggesting that mast cells are a major source of PGD₂ in asthma patients. These cells are most likely intraepithelial mast cells, previously shown to accumulate in TH₂-high asthmatic airway epithelium [30], and reported to increase [31] with disease severity [18].

Asthma GWAS genes show cell-type restricted expression (Fig. 4e). When excluding the widely expressed HLA genes from the analysis, fibroblasts and T cells express the highest number of asthma GWAS genes (Fig. 4f), which are also mostly upregulated in asthma (Fig. 4f). GATA3 expression is restricted to T cells (Fig. 4f) and increased in patients with asthma (Fig. 4f). We detected up-regulation of CD4 (but not CD8A) in the T cell cluster, suggesting an increase in TH₂ CD₄⁺ T cells (Fig. 4f). Therefore, we investigated the CD₄⁺ T cell compartment in airway wall biopsies in more detail.

Pathogenic effector T helper 2 cells are enriched in asthmatic airways

In line with the increased GATA3 and CD4 expression mentioned above, TH₂ cells are known to be key drivers of asthma [14,32]. To assess the presence of TH₂ effector cells in the airways of patients with asthma (Fig. 4f), we single-cell sorted CD₄⁺ T cells followed by in-depth transcriptional phenotyping by SmartSeq2 profiling (Fig. 5a and Methods). We analyzed cells from both peripheral blood and airway wall biopsies (Fig. 5a) from a larger cohort of patients with asthma and healthy controls (Fig. 5b and Supplementary Table 6). Unbiased clustering reveals six major populations of CD₄⁺ T cells (Fig. 5c and Extended Data Fig. 10) with no differences in their relative abundance between asthma and healthy airway wall (Fig. 5d).

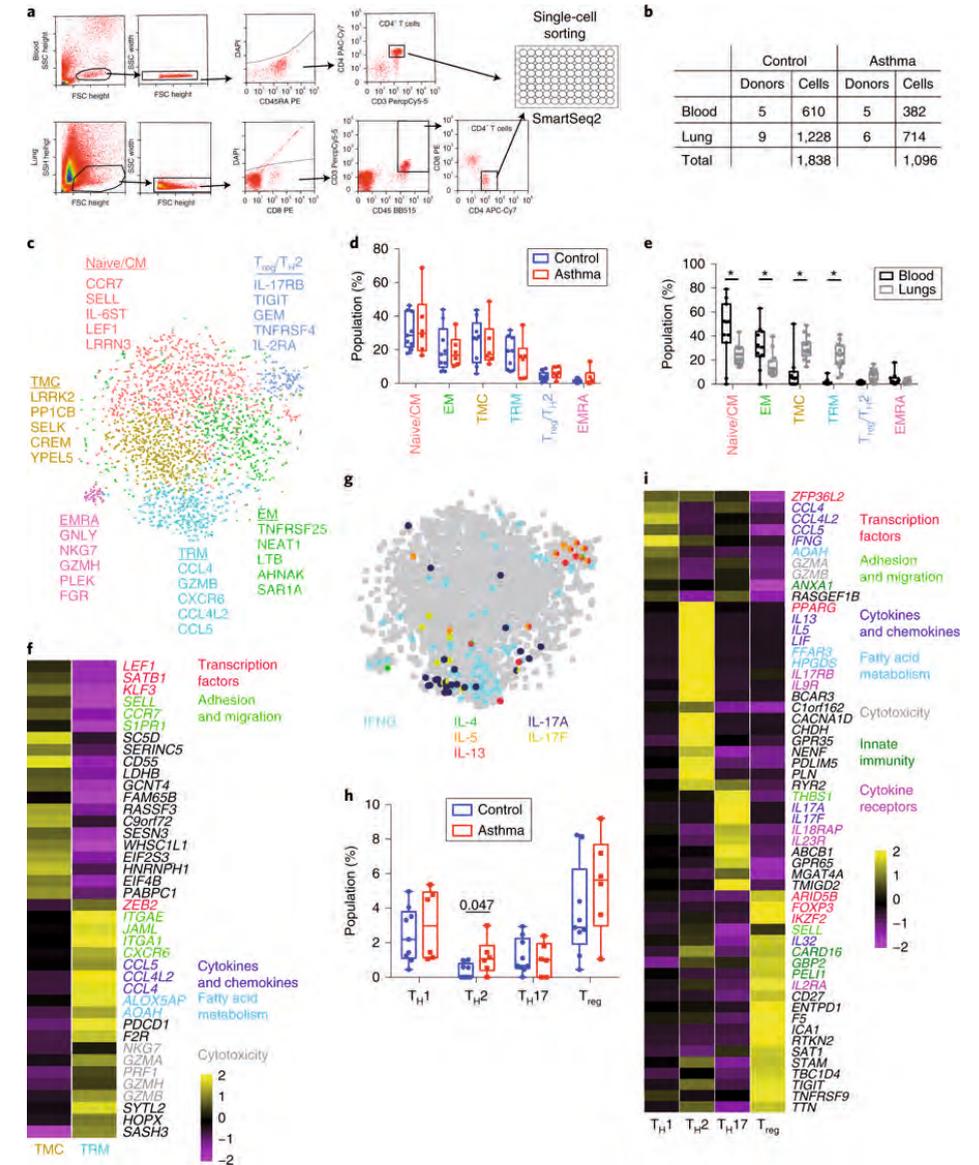


Figure 5: Pathogenic effector TH₂ cells are enriched in asthmatic airways, **a**: Schematic depicting experimental layout of single-cell sorting of CD₄⁺ T cells from blood and lung airway biopsies, **b**: Table with the number of donors by anatomical location for healthy control and asthma groups, **c**: t-SNE displaying clusters of T cells identified by analyzing the combined cells from blood and lung from healthy control and asthma groups, **d**: Box and whisker plots showing the cluster cell distributions from healthy control patients and patients with asthma, **e**: Box and whisker plots depicting the cluster composition per donor according to the tissue source from which the cells were isolated, **f**: Heat map showing the average expression per cluster of genes differentially expressed between the two lung specific CD₄⁺ T cell populations. Gene names colored according to functional categories, **g**: t-SNE depicting canonical cytokines from TH₁, TH₂, and TH₁₇ cells,

h: Box and whisker plots showing the number of TH1, TH2, and TH17 cells defined by canonical cytokines expression and T_{reg} cells identified by unbiased clustering, **i:** Heat map of average cluster gene expression of markers differentially expressed between TH1, TH2, TH17, and T_{reg} cells. Gene names colored according to functional categories. In **d**, **e** and **h**, box and whisker plots show all data points, and the box represents the second and third quartiles and median. Significance was analyzed using unpaired multiple t-tests assuming no consistent s.d. and corrected for multiple comparison using the Holm–Sidak method. All the differential expression analyses in **f** and **i** were performed using the non-parametric two-sided Wilcoxon rank sum test in Seurat. Panels **c–g** depict the number of cells and individuals described in **b**. Panels **h** and **i** depict the number of individuals described in **b**.

Comparative analysis of CD4⁺ T cells isolated from paired blood and lung samples allowed us to differentiate between tissue-resident and circulating T cells in an unbiased way (Fig. 5e and Methods). We identified two CD4⁺ T cell subsets highly enriched in airway wall: the classical TRM CD4⁺ T cells, and a novel subset, which we named the tissue migratory CD4⁺ T cell (TMC) subset (Fig. 5e). Naive/central memory (CM), effector memory (EM), and effector memory reexpressing CD45RA (EMRA) CD4⁺ T cells, as well as a mixed regulatory T cell (Treg)/TH2 cluster, are either enriched in blood or present in both blood and airway wall biopsies (Fig. 5e).

To better understand the two distinct lung airway-resident CD4⁺ T cell subsets, we performed differential expression analysis between TRM and TMC cells (Fig. 5f). TRM cells lack S1PR1 and CCR7 expression. TRM cells in airway wall also expressed high levels of CXCR6 and ITGAI1, chemokines (CCL4, CCL4L2, CCL5) and effector molecules (PRF1, GZMB, GZMA, GZMH) (Fig. 5f and Extended Data Fig. 10), indicating they are in a primed state capable of direct effector function, as recently shown for TRMs from lung parenchyma [33]. TMC cells expressed the tissue egression markers S1PR1, CCR7, and SELL (CD62L) (Fig. 5f) as well as several transcription factors highly expressed in circulating cells such as LEF1, SATB1, and KLF3. Small numbers of TMC cells were present in peripheral blood CD4⁺ T cells (Fig. 5e), suggesting that these cells might have the potential to transit between lung, lymph, and blood.

CD4⁺ effector T cells are classically divided into distinct functional subsets based on their cytokine profile [14]. We manually annotated clusters of TH1 (IFNG⁺), TH2 (IL4⁺, IL5⁺ or IL13⁺), and TH17 (IL17A⁺ or IL17F⁺) cells based on their cytokine expression profiles (Fig. 5g and Extended Data Fig. 10). Although rare overall (Supplementary Table 6), TH2 cells were significantly increased in the airway wall in patients with asthma, with no difference in the relative proportions of the other TH subsets (Fig.

5h). In addition to the signature cytokines IL4, IL5, and IL13 and the transcription factor GATA3 (Extended Data Fig. 10), airway wall TH2 cells express HGPDS, identifying them as pathogenic effector TH2 (peTH2) cells, previously associated with eosinophilic inflammation of the gastrointestinal tract and skin [34]. Airway TH2 cells also express the transcription factor PPARG and the cytokine receptors IL17RB and IL9R (Fig. 5i). IL17RB has been reported to be upregulated in pathogenic allergen-specific TH2 cells (coined TH2A cells), which are present in allergic disease [35] as well as in chronic rhinosinusitis with nasal polyps [36], suggesting airway wall TH2 cells share features with both TH2A and peTH2 cells.

Asthma is characterized by specific signaling networks

Asthma is characterized by remodeling of the airways, which depends on complex interactions between structural and inflammatory cells [14], both via direct physical interactions and secreted proteins and small molecules. We used our recently developed receptor/ligand database and statistical inference framework CellPhoneDB [37] (www.cellphonedb.org) to identify potential cell–cell interactions in the airway wall, and define their changes in asthma. While most interactions are unchanged, some were specific to the diseased or healthy states (Supplementary Table 7).

In healthy controls, the cell–cell interaction landscape of the airway wall was dominated by lung structural cells (mesenchymal and epithelial cell types) communicating with each other, and with both TRM and TMC CD4⁺ T cells (Fig. 6a,b, left panels). In the asthmatic airway wall, the number of predicted interactions between epithelial and mesenchymal cells was strongly reduced. Instead, the cell–cell communication landscape is dominated by TH2 cells in asthma. TH2 cells have increased predicted interactions with other immune cells, epithelial cells, and especially mesenchymal cells, both fibroblasts and smooth muscle cells (Fig. 6a,b, right panels).

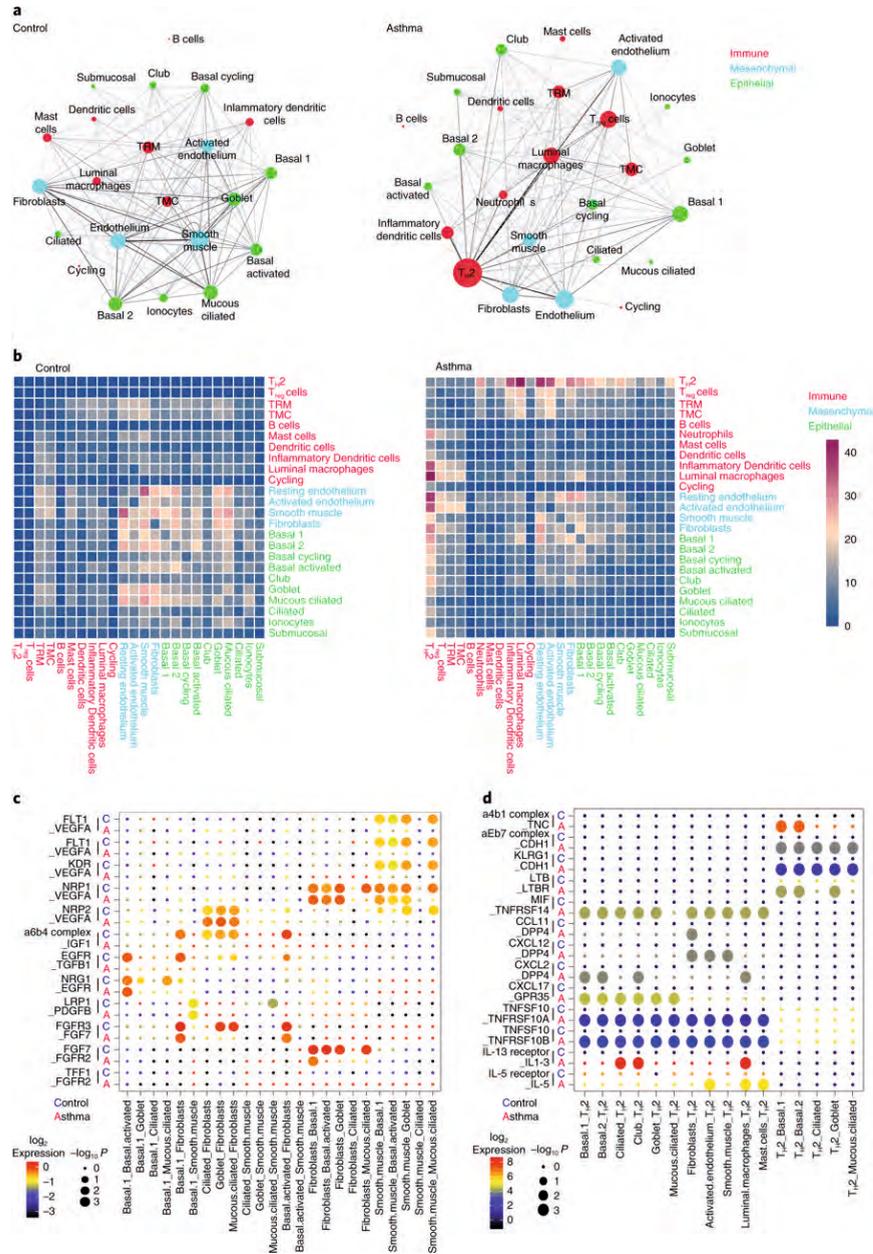


Figure 6: asthma is characterized by unique cell-to-cell signaling networks. We quantified the predicted cell interactions in healthy and asthmatic airways between all the epithelial and non-epithelial cell clusters identified in Figs. 3 and 4, plus the lung airway-enriched populations of CD4 T cells (TH₂, T_{H2}^{reg}, TMC, and TRM). **a:** Networks depicting cell types as nodes and interactions as edges. Size of cell type is proportional to the total number of interactions of each cell type, and edge thickness is proportional to the number of interactions between the connecting types. **b:** Heat map depicting the number of all possible interactions between the clusters analyzed. Cell types grouped by broad lineage (epithelia, mesenchymal, or immune). **c:** Dot plot depicting selected epithelial–epithelial and epithelial–mesenchymal interactions enriched in healthy airways but absent in asthmatic airways. **d:** Dot plot depicting selected epithelial–immune and mesenchymal–immune interactions highly enriched in asthmatic airways but absent in healthy airways. Panels a–d depict nine healthy individuals and six individuals with asthma, as shown in Fig. 5b.

Analysis of the predicted cell–cell interactions between structural cells in healthy airway wall revealed a wealth of growth factor signalling pathways including the fibroblast growth factor (FGF), epidermal growth factor receptor (EGFR), insulin-like growth factor (IGF), transforming growth factor (TGF), platelet-derived growth factor (PDGF) and vascular endothelial growth factor (VEGF) pathways (Supplementary Table 7). Cell–cell interactions unique to asthma included TH₂–epithelial cell contacts (for example, KLRG1 and CD103/E-cadherin; integrin/tenascin-C). Epithelial expression of alarmins and cytokines, such as IL-33, TSLP, and TNFSF10/TRAIL (Fig. 6d), all of which are known to play a role in asthma [38–40], might activate TH₂ cells expressing their receptors.

In addition to validating these well-known interactions, which for IL-33 and TSLP failed to reach statistical significance in our unbiased cell–cell interaction analysis, we identified novel epithelial–TH₂ cell interactions in asthma: the interactions between epithelial chemokines CXCL2, CXCL17, and their TH₂-expressed receptors (Fig. 6d). Interestingly, mesenchymal cells share some of these interactions, such as expression of TNFSF10/TRAIL and MIF. Predicted mesenchymal–TH₂ cell interactions in asthma are CXCL12 and CCL11, expressed by fibroblasts and smooth muscle cells. Airway wall TH₂ cells in asthma express IL-5 and IL-13 (Fig. 5j), with receptor expression by immune cells and epithelial cells, respectively, in line with the observed IL-13-driven gene signature in mucous ciliated and goblet cells in asthma (Extended Data Fig. 7). In addition to these classical TH₂ cytokines, TH₂ cells express LTB for which basal epithelial cells express the receptor.

Discussion

We describe the cellular landscape of human lung tissue at the single-cell level, charting differences in frequencies and molecular state of lung structural and inflammatory cells between upper and lower airways and parenchyma. We provide a first detailed molecular description of two separate tissue-resident subsets of CD4⁺ T cells in airway wall, one of which was hitherto unknown. We also conclusively show the presence of the recently identified [34–36] pathogenic effector TH2 cells in the airway wall in asthma, as evidenced by the combined expression of IL5, IL13, HPGDS, PPARG, and IL17RB.

We identify a novel mucous ciliated cell state in asthmatic airway epithelium that contributes to mucous cell metaplasia. The goblet-like gene expression profile in FOXJ1⁺ cells with a full-blown ciliated cell transcriptional phenotype strongly indicates that this molecular state is induced in ciliated cells by type 2 cytokines. Mucous metaplasia of ciliated cells and goblet cell hyperplasia both contribute to the increase in mucin-producing cells in asthma.

These changes in airway epithelium in asthma differ from those in patients with chronic rhinosinusitis with polyps [24]. In this type 2 inflammatory upper airway disease, IL-4/IL-13-driven gene transcription was observed in basal cells, which were arrested in differentiation and increased in frequency [24]. In asthma, basal cell numbers are not strongly increased. Instead, we observe expression of the IL-4/IL-13-driven gene signature mainly in goblet and mucous ciliated cells (Extended Data Fig. 7). Hence, chronic type 2 inflammation has divergent effects on the epithelia of the upper versus the lower airways. In contrast, the changes in the eicosanoid pathway observed in chronic type 2 inflammation of the upper [24] and lower (Extended Data Fig. 8) airways are very similar, underscoring the presence of common cellular mechanisms between these two anatomical locations.

Finally, comprehensive analysis of the cell–cell interactions in airway wall in asthma identifies dominance of TH2 cells interacting with structural and inflammatory cells. The extensive growth factor signaling between epithelial cells and mesenchymal cells observed in healthy airway wall is largely lost in asthmatic airway wall, at odds with a reactivation of the epithelial–mesenchymal trophic unit [41]. Instead, our data support a

shift in cellular phenotypes in airway wall due to the local production of TH2 cytokines in our patient cohort with mild to moderate childhood-onset asthma. This global view of the airway wall cellular landscape opens up new perspectives on lung biology and molecular mechanisms of asthma.

Methods

Patient recruitment and ethical approval

Bronchoscopy biopsy (10X and SmartSeq2 analysis)

Cohort inclusion criteria for all subjects were: age between 40 and 65 years old and a history of smoking <10 pack-years. For the patients with asthma, inclusion criteria were: age of onset of asthmatic symptoms ≤12 years old, documented history of asthma, use of inhaled corticosteroids with(out) β₂-agonists due to respiratory symptoms, and a positive provocation test (that is, PC₂₀ methacholine (concentration of methacholine needed to produce a 20% fall in the forced expiratory volume in the first second (FEV₁)) ≤ 8 mg ml⁻¹ with 2 min protocol). For the non-asthmatic controls, the following criteria were essential for inclusion: absent history of asthma, no use of asthma-related medication, a negative provocation test (that is PC₂₀ methacholine >8 mg ml, and adenosine 5'-monophosphate >320 mg ml with 2 min protocol), no pulmonary obstruction (that is, FEV₁/forced vital capacity (FVC) ≥ 70%) and absence of lung function impairment (that is FEV₁ ≥ 80% predicted).

Patients with asthma stopped inhaled corticosteroid use six weeks before all tests. All subjects were clinically characterized with pulmonary function and provocation tests, blood samples were drawn, and finally subjects underwent a bronchoscopy under sedation. If a subject developed upper respiratory symptoms, bronchoscopy was postponed for ≥6 weeks.

Fibreoptic bronchoscopy was performed using a standardized protocol during conscious sedation [42]. Six macroscopically adequate bronchial biopsies were collected for this study, located between the third and sixth generation of the right lower and middle lobe. Extracted biopsies were processed directly thereafter, with a maximum of 1 hour delay.

The medical ethics committee of the Groningen University Medical Center Groningen approved the study, and all subjects gave their written informed consent. Detailed patient information is given in Supplementary Table 3.

Lung resection (Drop-seq analysis)

Fresh resected human lung tissue (parenchymal lung and distal airway specimens) was obtained via the CPC BioArchive at the Comprehensive Pneumology Center Munich (CPC-M). In total, we analyzed parenchymal tissue of uninvolved areas of tumor resection material from four patients. All participants gave written informed consent and the study was approved by the local ethics committee of the Ludwig-Maximilians University of Munich.

For transport from the surgeon to the laboratory, lung tissue samples were stored in ice-cold DMEM-F12 media and packed in thermo stable boxes. Tissue was processed with a maximum delay of 2 hours after surgery. On delivery to the lab, tissue samples were assessed visually for qualification for the study.

Donor information is given in Supplementary Table 8.

Lung transplant tissue (10X analysis)

Human lung tissue was obtained from deceased organ donors from whom organs were being retrieved for transplantation. Informed consent for the use of tissue was obtained from the donors' families (REC reference: 15/EE/0152 NRES Committee East of England—Cambridge South).

Fresh tissue from the peripheral parenchyma of the left lower lobe or lower right lobe of the lung was excised within 60 min of circulatory arrest and preserved in University of Wisconsin (UW) organ preservation solution (Belzer UW Cold Storage Solution, Bridge to Life) until processing.

Donor 284C. Gender: male. Age band: 55–60. BMI: 25.83. Cause of death: hypoxic brain damage. Smoking history: smoked 20 per day for 25 yr. Stopped: 08/2000. Respiratory related information: chest X-ray normal on admission. No pleural effusion or pneumothorax. Not diagnosed with asthma, but inhalers for possible seasonal

wheeze. Family report only using inhaler approximately five times per year. No recent peak flow on record last one in 2008 when it was 460, predicted is 611. Time from death to cell lysis: 12 h.

Donor 290B. Gender: female. Age band: 60–65. BMI: 27.77. Cause of death: hypoxic brain damage. Smoking history: smoked 15 per day for 7 yr. Stopped: no details. Respiratory related information: respiratory tests all normal on admission; maintaining own airway. GP notes report acute bronchitis in 1994. Time from death to cell lysis: 2 h 27 min.

Donor 292B. Gender: male. Age band: 55–60. BMI: 27.44. Cause of death: intracranial hemorrhage. Smoking history: smoked 20 per day for 46 yr. Stopped: no details. Respiratory related information: chest X-ray normal on admission, lungs appear clear. Bronchoscopy results show global inflamed mucosa. No other history of respiratory issues. Time from death to cell lysis: 18 h 50 min,

Donor 296C. Gender: female. Age band: 30–35. BMI: 20.9. Cause of death: intracranial hemorrhage. Smoking history: smoked 20 per day for 19 yr. Stopped: no details. Respiratory related information: chest X-ray shows collapsed left lobe on admission due to consolidation. Right lobe looks normal. No history or record of respiratory issues. Time from death to cell lysis: 15 h 30 min.

Donor 298C. Gender: male. Age band: 50–55. BMI: 24. Cause of death: intracranial hemorrhage. Smoking history: not available. Stopped: no details. Respiratory related information: no details. Time from death to cell lysis: 15 h 30 min.

Donor: 302C. Gender: male. Age band: 40–45. BMI: 34.33. Cause of death: known or suspected suicide. Smoking history: smoked 20 per day for 25 yr. Stopped: no details. Respiratory related information: chest X-ray shows reduced volume in right lung due to collapsed right lower lobe on admission. No history or record of respiratory issues. Time from death to cell lysis: 13 h 30 min.

Archived formalin-fixed paraffin-embedded lung blocks

Left-over frozen peripheral lung tissues from six current smokers and four non-smokers who underwent lung resection surgery. These subjects did not have a history of lung

disease, apart from lung cancer for which the patients underwent surgery. Lung tissue samples were taken as distant from the tumor as possible. Thus, any possible effect of the tumor on the lung tissue was minimized. All samples were obtained according to national and local ethical guidelines and the research code of the University Medical Center Groningen. Sample information is given in Supplementary Table 9.

Blood processing

Lithium heparin-anticoagulated whole blood (500 μ l) was lysed using an ammonium chloride-potassium solution (155 mM ammonium chloride (NH_4Cl), 10 mM potassium bicarbonate (KHCO_3), 0.1 mM EDTA). Cells were centrifuged for 5 min at 4 °C, 550g, after which the cell pellet was washed twice with PBS containing 1% BSA, followed by staining for cell surface markers.

Lung tissue processing

Bronchoscopy biopsy

A single-cell solution was obtained by chopping the biopsies finely using a single edge razor blade. The chopped tissue was then put in a mixture of 1 mg ml⁻¹ collagenase D and 0.1 mg ml⁻¹ DNase I (Roche) in HBSS (Lonza). This was then placed at 37 °C for 1 h with gentle agitation. The single-cell suspension was forced through a 70 μ m nylon cell strainer (Falcon). The suspension was centrifuged at 550g, 4 °C for 5 min and washed once with a PBS containing 1% BSA (Sigma-Aldrich). The single-cell suspensions used for 10x Genomics scRNA-seq analysis were cleared of red blood cells by using a red blood cell lysis buffer (eBioscience) followed by staining for cell surface markers.

Lung tissue resection

For each sample, 1.0–1.5 g of tissue was homogenized by mincing with scissors into smaller pieces (~0.5 mm² per piece). Before tissue digestion, lung homogenates were cleared from excessive blood by addition of 35 ml of ice-cold PBS, followed by gentle shaking and tissue collection using a 40 μ m strainer. The bloody filtrate was discarded. The tissue was transferred into 8 ml of enzyme mix consisting of dispase (50 caseolytic U ml⁻¹), collagenase (2 mg ml⁻¹), elastase (1 mg ml⁻¹), and DNase (30 μ g ml⁻¹) for mild enzymatic digestion for 1 h at 37 °C while shaking. Enzyme activity was inhibited by adding 5 ml of PBS supplemented with 10% FCS. Dissociated cells in suspension were passed through a 70 μ m strainer and centrifuged at 300g for 5 min at 4 °C. The cell pellet

was resuspended in 3 ml of red blood cell lysis buffer and incubated at room temperature for 2 min to lyse remaining red blood cells. After incubation, 10 ml of PBS supplemented with 10% FCS was added to the suspension and the mix was centrifuged at 300g for 5 min at 4 °C. The cells were taken up in 1 ml of PBS supplemented with 10% FCS, counted using a Neubauer chamber, and critically assessed for single-cell separation. Dead cells were counted to calculate the overall cell viability, which needed to be above 85% to continue with Drop-seq. Two-hundred and fifty thousand cells were aliquoted in 2.5 ml of PBS supplemented with 0.04% of bovine serum albumin and loaded for Drop-seq at a final concentration of 100 cells μ l⁻¹.

Rejected lung transplant

For each sample, 1–2 g of tissue was divided in 5 smaller pieces then transferred to 5 ml eppendorfs containing 1.5 ml 0.5 mg ml⁻¹ collagenase D and 0.1 mg ml⁻¹ DNase I (Sigma) in RPMI. Samples were then finely minced using scissors. Minced tissue was then transferred to a Petri dish and extra digestion medium added to completely cover the tissue. Samples were incubated 30 min at 37 °C. Cells were then passed up and down through a 16-gauge needle 10 times. Samples were incubated for an additional 15 min at 37 °C. Cells were filtered a 70 μ m filter, then spun down for 6 min 1,400 r.p.m. One milliliter of red blood cell lysis (eBioscience) was added to the pellet during 5 min. Cells were resuspended in RPMI + 10%FCS and counted. Dead cells were removed using the Dead Cell Removal Kit (Miltenyi Biotec). In brief, cells were incubated with anti-annexin V beads for 15 min. The cell suspension was then passed through a magnetic column and dead annexin V⁺ cells remained in the column, while live cells were collected. Viability was then estimated via trypan blue. More than 99% of cells were viable.

Flow cytometry

Blood leukocytes were stained with CD4 APC-Cy7, CD3 PerCP Cy5.5, CD8 APC, and CD45RA-PE (eBioscience) for 30 min at 4 °C and washed twice with PBS containing 1% BSA. Propidium iodide was added 5 min before sorting.

Airway wall biopsy single-cell suspensions were stained for 30 min at 4 °C with CD3 PerCP Cy5.5, CD45 BB515, CD4 APC-Cy7 (BD), and CD8 PE and washed twice with PBS containing 1% BSA. Propidium iodide was added 5 min before sorting.

Cell sorting

Lymphocytes were selected in the FCS/SSC plot. These were then selected on single, live cells for blood or single, live, CD45⁺ for lung. The sorted cells were positive for CD3 and CD4 as shown in Fig. 5a. All cells were sorted in a MoFlo Astrios (Beckman Coulter) using Summit Software (Beckman Coulter).

Immunohistochemical staining

Human lung tissue containing large airways were collected from archival formalin-fixed paraffin-embedded blocks (n=10, 6 smokers and 4 non-smokers). Serial sections (~4 µm) were cut for immunohistochemistry (IHC) and immunofluorescent (IF) staining.

Serial sections from formalin-fixed paraffin-embedded lung tissue were stained for using standard protocols, with antibodies specified in the figures. Briefly, serial sections were deparaffinized in xylene, rehydrated, and immersed in 10 mM sodium citrate buffer (pH 6.0). Antigen retrieval was performed by boiling the sections in a pressure cooker at 120 °C for 20 min.

IHC and IF staining was performed as described previously [42,43]. For the IHC staining cells were stained with a primary antibody (see below for antibody details) and visualized with diaminobenzidine (DAB, Sigma) solution. For the IF staining, cells were stained with primary antibody. Secondary antibodies conjugated to fluorophores (donkey anti-rabbit-488, donkey anti-mouse-555) were used at a dilution of 1:100. DAPI, dissolved in Dako Fluorescence Mounting Medium (Dako S3023) at a dilution of 1:1,000, was used as a nuclear stain.

Chromium 10X Genomics library and sequencing

Airway biopsy

Single-cell suspensions were manually counted using a haemocytometer and concentration adjusted to a minimum of 300 cells µl⁻¹. Cells were loaded according to standard protocol of the Chromium single-cell 3' kit to capture between 2,000 and 5,000 cells per chip position. All the following steps were performed according to the standard protocol. Initially, we used one lane of an Illumina HiSeq 4000 per 10x Genomics chip

position. Additional sequencing was performed to obtain coverage of at least mean coverage of 100,000 reads per cell.

Lung transplant

Single-cell suspensions were manually counted using a haemocytometer and concentration adjusted to 1,000 cells µl⁻¹. Cells were loaded according to standard protocol of the Chromium single-cell 3' kit to capture between 2,000 and 5,000 cells per chip position. All the following steps were performed according to the standard manufacturer protocol. Initially, we used one lane of an Illumina HiSeq 4000 per 10x Genomics chip position. Additional sequencing was performed to obtain coverage of at least mean coverage of 100,000 reads per cell.

Antibody list

A full antibody list is given in Supplementary Table 10.

SmartSeq 2 library preparation and sequencing

Library preparation was performed with minor modifications from the published SmartSeq2 protocol [44]. In short, single cells were flow sorted onto individual wells of 96 or 384 wells containing 4 µl (96 wells) or 1 µl (384 wells) of lysis buffer (0.3% triton plus DNTPs and OligoDT). After sorting, plates were frozen and stored at -80 °C until further processing. PCR with reverse transcription (25 cycles) and Nextera library preparation performed as described in ref. [44].

Drop-seq library preparation and sequencing

Drop-seq experiments were performed largely as described previously [2] with few adaptations during the single-cell library preparation. Briefly, using a microfluidic polydimethylsiloxane device (Nanoshift), single cells (100 µl⁻¹) from the lung cell suspension were co-encapsulated in droplets with barcoded beads (120 µl⁻¹, purchased from ChemGenes) at rates of 4,000 µl h⁻¹. Droplet emulsions were collected for 15 min each before droplet breakage by perfluorooctanol (Sigma-Aldrich). After breakage, beads were collected and the hybridized mRNA transcripts reverse transcribed (Maxima RT, Thermo Fisher). Unused primers were removed by the addition of exonuclease I (New England Biolabs), following which beads were washed, counted, and aliquoted for pre-amplification (2,000 beads per reaction, equals ~100 cells per reaction) with 12 PCR

cycles (primers, chemistry, and cycle conditions identical to those previously described). PCR products were pooled and purified twice by 0.6x clean-up beads (CleanNA). Before tagmentation, cDNA samples were loaded on a DNA High Sensitivity Chip on the 2100 Bioanalyzer (Agilent) to ensure transcript integrity, purity, and amount. For each sample, 1 ng of pre-amplified cDNA from an estimated 1,000 cells was tagmented by Nextera XT (Illumina) with a custom P5 primer (Integrated DNA Technologies). Single-cell libraries were sequenced in a 100 bp paired-end run on the Illumina HiSeq4000 using 0.2 nM denatured sample and 5% PhiX spike-in. For priming of read 1, 0.5 μ M Read1CustSeqB (primer sequence: GCCTGTCCGCGAAGCAGTGGTATCAACGCAGAGTAC) was used.

Bulk transcriptome

Biopsies were fresh frozen in liquid nitrogen and stored at -80°C . RNA was extracted after a few weeks using a combination of Trizol and the RNeasy MinElute Clean Up kit from Qiagen. RNA was prepared from sequencing using the TruSeq RNA Library Prep Kit v2. Samples were then sequenced in a HiSeq 4000.

Single-cell RNA sequencing data alignment

For SmartSeq2 raw sequencing data, paired-end reads were mapped to the human genome (GRCh38) using GSNAP with default parameters [45]. Then, uniquely mapped reads were counted using htseq-count (<http://www-huber.embl.de/users/anders/HTSeq/>). Low-quality cells were filtered out using the outlier detection algorithm in R Scater package based on a cut-off of $2 \times$ median absolute deviation.

10x Genomics raw sequencing data were processed using CellRanger software version 2.0.2 and the 10x human genome GRCh38 1.2.0 release as the reference.

The Drop-seq core computational pipeline was used for processing next generation sequencing reads of the Drop-seq scRNA-seq data, as previously described [2]. Briefly, STAR (version 2.5.2a) was used for mapping [46]. Reads were aligned to the human reference genome hg19 (provided by Drop-seq group, GSE63269).

Bulk transcriptome computational analysis

The bulk samples were aligned using STAR 2.5.1b, using the STAR index from the GRCh38 reference that was used when mapping 10x data, and quantified using HTSeq.

The data were then processed using the Seurat-inspired workflow within Scanpy, adding a number of ‘pseudo-bulks’ obtained by taking 10x data from donors matching the bulk samples and summing expression across all cells.

Data quality control

General strategy for 10x datasets

Optimal tissue dissociation conditions are cell-type dependent, resulting in a certain degree of cell lysis when working with a mixed tissue sample. This results in substantial background levels of ambient RNA in the single-cell suspension that vary with cell-type composition, so we applied SoupX for background correction (see below). We analyzed each donor sample separately and excluded cells with a number of genes higher than the median + 2 s.d. for that donor. We further excluded cell with high number of unique molecular identifiers (UMIs) and high percentage of mitochondrial reads (see below).

In parallel, we used scrublet (see below) to infer the number of the doublets in the dataset before applying the filters previously described and excluded any remaining cells predicted to be doublets that were still present in the dataset. We normalized and scaled our data (see below), performed clustering (see below), and identified and subset the data into epithelial and non-epithelial cell groups (as shown in Extended Data Figs 1 and 5). After separation between epithelial and non-epithelial, we clustered the cells and performed curated doublet removal (see below) based on known lineage restricted markers.

General strategy for Drop-seq data

We normalized and scaled the data, then performed filtering based on the number of genes and percentage of mitochondrial reads.

General strategy SmartSeq2 data

We normalized and scaled the data, then performed filtering based on the number of genes and percentage of mitochondrial reads. To avoid potential batch effects from the lung digestion protocol, we corrected the gene expression of the CD4 SmartSeq2 dataset using a small subset of genes, the expression of which has been recently shown to be highly responsive to enzymatic digestions [47]: FOS, ZFP36, JUN, FOSB, HSPA1A, JUNB, EGR1, UBC.

Ambient RNA correction (SoupX)

Different batches can be affected by different levels of ambient RNA. To take this into account, we used the recently developed SoupX method [48]. Briefly, ambient RNA expression is estimated from the empty droplet pool (to UMI or less). Expression of these genes in each cell is then calculated and compared with their proportion in the ambient RNA profile. Transcripts with a bimodal profile (that is, that characterize specific groups of cells but are also highly abundant in empty droplets) are then grouped based on their function. The contamination fraction derived from the expression of these genes is then used to calculate the fraction of each droplet's expression corresponding to the actual cell. Finally, this fraction and the ambient profiles are subtracted from the real expression values.

UMI and number of genes filtering

10x data (after SoupX correction)

nUMI: minimum 1000/maximum 60000. percent.mito, minimum 0/maximum = 3%.

SmartSeq2 data

nGene: minimum 1000/maximum 4000. percent.mito, minimum 0/maximum = 15%.

Drop-seq data

nGene: minimum 200/maximum 4000. percent.mito, minimum 0/maximum = 20%.

Scrublet

We used Scrublet [49] for unbiased computational doublet inference. Doublets were identified in each 10x sample individually using Scrublet, setting the expected doublet rate to 0.03 and keeping all other parameters at their default values. Cells were excluded when they had a score higher than 0.1 for upper and lower airway samples or higher than 0.05 for parenchyma samples.

Normalization and scaling

Downstream analyses including, normalization, scaling, clustering of cells, and identifying cluster marker genes were performed using the R software package Seurat [50] version 2.1 (<https://github.com/satijalab/seurat>).

Samples were log normalized and scaled for the number of genes, number of UMIs, and percentage of mitochondrial reads. The epithelial biopsy dataset comparing healthy and asthma was also scaled for XIST expression, as we observed some gender specific clusters of cells that shared lineage markers with the other observed clusters.

Curated doublet removal

In addition to the general quality control described above, we combined literature knowledge about cell lineages with computational clustering to identify clusters enriched in potential doublets. The strategy for each dataset is described below.

Lung atlas epithelial dataset

We removed cells with expression level higher than 0.5 for any of the following markers: PTPRC (immune), FCER1G (immune), PDGFRA (fibroblast), or PECAM1 (endothelial) (Fig. 1 and associated Extended Data figures).

Lung atlas non-epithelial dataset

We removed cells with expression level higher than 0.5 for any of the following markers: EPCAM (epithelial), KRT5 (basal), FOXJ1 (ciliated), or MUC5AC (secretory). We then performed first clustering round (7 PCs, resolution 2) and excluded clusters that expressed combinations of the following lineage specific markers: MARCO (macrophage), CCL21 (lymphatic endothelial), TPSB2 (mast cell), or CD3D (T cell). We performed a second clustering round and exclude a cluster formed by cells from one donor that had low expression TPSB2, while lacking markers for all other immune lineages (Fig. 2 and associated Extended Data figures).

Asthma biopsy epithelial cells

Owing to the smaller number of cells, in addition to the general quality control metrics, we only performed cluster-based doublet exclusion, without cell filtering. We performed one round of clustering and removed one clusters with high expression of PECAM1 (endothelial marker) (Fig. 3 and associated Extended Data figures).

Asthma biopsy non-epithelial cells

In addition to the general quality control metrics, we performed three rounds of clustering where we excluded clusters with high levels of EPCAM or KRT5 expressed

in much higher levels than immune lineage markers (Fig. 4 and associated Extended Data figures).

Dimensionality reduction

We performed principal component analysis (PCA) dimensionality reduction with the highly variable genes as input. We then used the PCs to calculate t-distributed stochastic neighbor embedding (t-SNE) for each dataset, using a perplexity value of 50.

Data clustering

We used the function ‘FindClusters’ from Seurat. In brief, this method uses a shared nearest neighbor (SNN) modularity optimization-based clustering algorithm to identify clusters of cells based on their PCs. Before constructing the SNN graph, this function calculates *k*-nearest neighbors (we used *k* = 30) and then it constructs the SNN graph. The number of PCs used for each clustering round was dataset dependent and they were estimated by the elbow of a PCA scree plot, in combination with manual exploration of the top genes from each PC.

matchScore

We used matchScore [51] to quantify the overlap of cell-type marker signatures between experiments, which is based on the Jaccard index. Only marker genes with adjusted *P* value < 0.1 and average log fold change > 1 were considered.

CellPhoneDB

We developed a manually curated repository of ligands, receptors and their interactions called CellPhoneDB (www.cellphonedb.org) [37], integrated with a statistical framework for predicting cell–cell communication networks from single-cell transcriptome data. Briefly, the method infers potential receptor–ligand interactions based on expression of a receptor by one cell type and a ligand by another cell type. Only receptors and ligands expressed in more than 30% of the cells in the specific cluster were considered. To identify the most relevant interactions between cell types, the method prioritizes ligand–receptor interactions that have cell type-specific expression. To this end, pairwise cluster–cluster interaction analyses were performed by randomly permuting the cluster labels of each cell 1,000 times. For each permutation, the total mean of the average receptor expression level of a cluster and the average ligand expression level of

the interacting cluster is calculated, and a null distribution is derived for each receptor–ligand pair in each cluster–cluster interaction. An empirical *P* value is calculated from the proportion of the means which are ‘as or more extreme’ than the actual mean. For the multi-subunit heteromeric complexes, the member of the complex with the minimum average expression is used for calculating the mean.

Network visualization was done using Cytoscape (version 3.5.1). All the interaction pairs with collagens were removed from the analysis. The networks layout was set to force-directed layout.

Trajectory analysis

Trajectory analysis was performed using Monocle version 2.2.0 (ref. 20). We ordered the cells onto a pseudotime trajectory based on the union of highly variable genes obtained from all cells, as well as those from only healthy or asthmatic donors.

Supervised analyses using GWAS genes

Asthma-associated GWAS gene list was collected using the GWAS Catalog of EMBL-EBI searching for the term asthma (<https://www.ebi.ac.uk/gwas/>). The list was downloaded on 8 February 2018. We took the genes that are in the top 50 hits of our single-cell differential expression marker list (either epithelial or non-epithelial) and asthma-associated GWAS list (the ‘matched’ gene list). We then hierarchically clustered the expression matrix of the matched gene list along its rows (genes) and columns (single cells) and represented this as a heat map.

Neuroendocrine cell identification

Neuroendocrine cells were identified by the expression of *CHGA*. Any cell expressing any amount of *CHGA* was classified as a neuroendocrine cell.

OMIM search for lung diseases

We searched the clinical synopses with known molecular basis in the OMIM database for the following terms: ‘pulm*’ or ‘bronchi*’ or ‘alveol*’ or ‘surfactant’ and retrieved 337 entries. These terms were chosen to minimize the return of genetic conditions causing respiratory insufficiency as a consequence of neuromuscular dysfunction, skeletal dysplasia (small rib cage), or lung segmentation defects arising in early embryogenesis.

These 337 entries were then manually curated to identify those conditions with features affecting the bronchial tree, alveoli, lung parenchyma, and pulmonary vasculature. On manual review, entries containing terms such as ‘alveolar ridge’ of the jaw and ‘pulmonary valve stenosis’ and ‘pulmonary embolism’, but no terms related to primary pulmonary disorders, were excluded from further consideration. Syndromes caused by chromosomal disorder or contiguous gene deletion were excluded.

Cluster specific marker expression were generated by comparing this list to the genes present in the top 50 (per cluster) of our list of differentially expressed genes.

Statistical methods

For 10x samples comparing healthy versus asthma, we used Fisher’s exact test corrected for multiple testing with Bonferroni method. Normalized CD4 cluster proportions (percentage of total cells) were analyzed using unpaired multiple t-tests assuming no consistent s.d. and corrected for multiple comparison using the Holm–Sidak method. We used a non-parametric two-sided Wilcoxon rank sum test in Seurat to identify differentially expressed genes in all the comparisons discussed.

Data availability

Data requests for raw and analyzed data and materials will fall under two categories. Datasets from healthy live volunteers and live volunteers with asthma will be promptly reviewed by the University of Groningen. Any data and materials that can be shared will be released via a Material Transfer Agreement. These datasets can be found on European Genome–phenome Archive (<https://www.ebi.ac.uk/ega/home>) EGAS00001001755. Datasets generated from lung resection samples using Drop-seq can be accessed in GSE130148. Datasets generated from deceased donors fall under Open Access Policies of the Human Cell Atlas (<https://www.humancellatlas.org> for details). This data can be accessed at European Genome–phenome Archive (<https://www.ebi.ac.uk/ega/home>) EGAS00001002649. Interactive exploration tool: www.lungcellatlas.org.

References

- 1 Tata PR & Rajagopal J. Plasticity in the lung: making and breaking cell identity. *Development* 144, 755–766 (2017).
- 2 Macosko EZ, *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214 (2015).
- 3 Montoro DT, *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 560, 319–324 (2018).
- 4 Plasschaert LW, *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 560, 377–381 (2018).
- 5 Bisse LR & Schmid-Grendelmeier P. Chemokines and their receptors in the pathogenesis of allergic asthma: progress and perspective. *Curr. Opin. Pulm. Med.* 11, 35–42 (2005).
- 6 Colvin RA, *et al.* Synaptotagmin-mediated vesicle fusion regulates cell migration. *Nat. Immunol.* 11, 495–502 (2010).
- 7 Urawa M, *et al.* Protein S is protective in pulmonary fibrosis. *J. Thromb. Haemost.* 14, 1588–1599 (2016).
- 8 Wujak A, *et al.* FXYD1 negatively regulates Na⁺/K⁺-ATPase activity in lung alveolar epithelial cells. *Respir. Physiol. Neurobiol.* 220, 54–61 (2016).
- 9 Krotova K, *et al.* Alpha-1 antitrypsin-deficient macrophages have increased matriptase-mediated proteolytic activity. *Am. J. Respir. Cell Mol. Biol.* 57, 238–247 (2017).
- 10 Vogl T, *et al.* S100A12 is expressed exclusively by granulocytes and acts independently from MRP8 and MRP14. *J. Biol. Chem.* 274, 25291–25296 (1999).
- 11 Mitchell A, *et al.* LILRA5 is expressed by synovial tissue macrophages in rheumatoid arthritis, selectively induces pro-inflammatory cytokines and IL-10 and is regulated by TNF- α , IL-10 and IFN- γ . *Eur. J. Immunol.* 38, 3459–3473 (2008).
- 12 Condon TV, Sawyer, RT, Fenton, MJ & Riches DWH. Lung dendritic cells at the innate-adaptive immune interface. *J. Leukoc. Biol.* 90, 883–895 (2011).
- 13 Baumann U, Routes, JM, Soler-Palacin P & Jolles S. The lung in primary immunodeficiencies: new concepts in infection and inflammation. *Front. Immunol.* 9, 1837 (2018).
- 14 Holgate ST, *et al.* Asthma. *Nat. Rev. Dis. Primers* 1, 15025 (2015).
- 15 Lopez-Guisa JM, *et al.* Airway epithelial cells from asthmatic children differentially express proremodeling factors. *J. Allergy Clin. Immunol.* 129, 990–997.e6 (2012).
- 16 Alcala SE, *et al.* Mitotic asynchrony induces transforming growth factor- β 1 secretion from airway epithelium. *Am. J. Respir. Cell Mol. Biol.* 51, 363–369 (2014).
- 17 Harkness, LM, Ashton AW & Burgess JK. Asthma is not only an airway disease, but also a vascular disease. *Pharmacol. Ther.* 148, 17–33 (2015).

- 18 Balzar S, et al. Mast cell phenotype, location, and activation in severe asthma. Data from the Severe Asthma Research Program. *Am. J. Respir. Crit. Care Med.* 183, 299–309 (2011).
- 19 Truyen E, et al. Evaluation of airway inflammation by quantitative TH1/TH2 cytokine mRNA measurement in sputum of asthma patients. *Thorax* 61, 202–208 (2006).
- 20 Trapnell C, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386 (2014).
- 21 Erle DJ & Sheppard D. The cell biology of asthma. *J. Cell Biol.* 205, 621–631 (2014).
- 22 Danahay H, et al. Notch2 is required for inflammatory cytokine-driven goblet cell metaplasia in the lung. *Cell Rep.* 10, 239–252 (2015).
- 23 Gomi K, Arbelaez V, Crystal RG & Walters MS. Activation of NOTCH1 or NOTCH3 signaling skews human airway basal cell differentiation toward a secretory pathway. *PLoS ONE* 10, e0116507 (2015).
- 24 Ordovas-Montanes, J et al. Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* 560, 649–654 (2018).
- 25 Luo W, et al. Airway epithelial expression quantitative trait loci reveal genes underlying asthma and other airway diseases. *Am. J. Respir. Cell Mol. Biol.* 54, 177–187 (2016).
- 26 Wu CA, et al. Bronchial epithelial cells produce IL-5: implications for local immune responses in the airways. *Cell. Immunol.* 264, 32–41 (2010).
- 27 Laitinen LA, Laitinen A & Haahtela T. Airway mucosal inflammation even in patients with newly diagnosed asthma. *Am. Rev. Respir. Dis.* 147, 697–704 (1993).
- 28 Arima M & Fukuda T. Prostaglandin D2 and TH2 inflammation in the pathogenesis of bronchial asthma. *Korean J. Intern. Med.* 26, 8–18 (2011).
- 29 Xue L, et al. Prostaglandin D2 activates group 2 innate lymphoid cells through chemoattractant receptor-homologous molecule expressed on TH2 cells. *J. Allergy Clin. Immunol.* 133, 1184–1194 (2014).
- 30 Dougherty RH, et al. Accumulation of intraepithelial mast cells with a unique protease phenotype in TH2-high asthma. *J. Allergy Clin. Immunol.* 125, 1046–1053.e8 (2010).
- 31 Hol BE, van de Graaf EA, Out TA, Hische EA & Jansen HM. IgM in the airways of asthma patients. *Int. Arch. Allergy Appl. Immunol.* 96, 12–18 (1991).
- 32 Muehling, LM, Lawrence, MG & Woodfolk JA. Pathogenic CD4⁺ T cells in patients with asthma. *J. Allergy Clin. Immunol.* 140, 1523–1540 (2017).
- 33 Oja AE, et al. Trigger-happy resident memory CD4⁺ T cells inhabit the human lungs. *Mucosal Immunol.* 11, 654–667 (2018).
- 34 Mitson-Salazar A, et al. Hematopoietic prostaglandin D synthase defines a proeosinophilic pathogenic effector human TH2 cell subpopulation with enhanced function. *J. Allergy Clin. Immunol.* 137, 907–918.e9 (2016).
- 35 Wambre E, et al. A phenotypically and functionally distinct human TH2 cell subpopulation is associated with allergic disorders. *Sci. Transl. Med.* 9, eaam9171 (2017).
- 36 Lam EPS, et al. IL-25/IL-33-responsive TH2 cells characterize nasal polyps with a default TH17 signature in nasal mucosa. *J. Allergy Clin. Immunol.* 137, 1514–1524 (2016).
- 37 Vento-Tormo R, et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* 563, 347–353 (2018).
- 38 Weckmann M, Kopp MV, Heinzmann A & Mattes J. Haplotypes covering the TNFSF10 gene are associated with bronchial asthma. *Pediatr. Allergy Immunol.* 22, 25–30 (2011).
- 39 Harada M, et al. Thymic stromal lymphopoietin gene promoter polymorphisms are associated with susceptibility to bronchial asthma. *Am. J. Respir. Cell Mol. Biol.* 44, 787–793 (2011).
- 40 Grotenboer NS, Ketelaar ME, Koppelman GH & Nawijn MC. Decoding asthma: translating genetic variation in IL33 and IL1RL1 into disease pathophysiology. *J. Allergy Clin. Immunol.* 131, 856–865 (2013).
- 41 Holgate ST, et al. Epithelial-mesenchymal communication in the pathogenesis of chronic asthma. *Proc. Am. Thorac. Soc.* 1, 93–98 (2004).
- 42 Heijink IH, et al. Down-regulation of E-cadherin in human bronchial epithelial cells leads to epidermal growth factor receptor-dependent TH2 cell-promoting activity. *J. Immunol.* 178, 7678–7685 (2007).
- 43 Song J, et al. Aberrant DNA methylation and expression of SPDEF and FOXA2 in airway epithelium of patients with COPD. *Clin. Epigenetics* 9, 42 (2017).
- 44 Picelli S, et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181 (2014).
- 45 Wu TD & Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26, 873–881 (2010).
- 46 Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).
- 47 van den Brink SC, et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* 14, 935–936 (2017).
- 48 Young MD, & Behjati S. SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data. Preprint at <https://doi.org/10.1101/303727> (2018).
- 49 Wolock SL., Lopez R & Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* 8, P281–291.E9 (2019).
- 50 Butler A, Hoffman P, Smibert P, Papalexi E & Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420 (2018).
- 51 Mereu E, et al. matchScore: matching single-cell phenotypes across tools and experiments. Preprint at <https://doi.org/10.1101/314831> (2018)