

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

A cellular census of healthy lung and asthmatic airway wall identifies novel cell states in health and disease

Vieira Braga, F.A.^{1,11*}, Kar, G.^{1,11*}, Berg, M.^{2,3,*}, Carpaij, O.A.^{3,4,§}, Polanski, K.^{1,§}, Simon, L.M.⁵, Brouwer, S.^{2,3}, Gomes, T.¹, Hesse, L.^{2,3}, Jiang, J.^{2,3}, Fasouli, E.S.^{1,11}, Efremova, M.¹, Vento-Tormo, R.¹, Affleck, K.⁷, Palit, S.⁵, Strzelecka, P.^{1,13,14}, Firth, H.V.¹, Mahbubani, K.T.A.⁶, Cvejic, A.^{1,13,14}, Meyer K.B.¹, Saeb-Parsy, K.⁶, Luinge, M.^{2,3}, Brandsma, C.-A.^{2,3}, Timens, W.^{2,3}, Angelidis, I.⁹, Strunz, M.⁹, Koppelman, G.H.^{3,10}, van Oosterhout, A.J.⁷, Schiller, H.B.⁹, Theis, F.J.^{5,8}, van den Berge, M.^{3,4}, Nawijn, M.C.^{2,3,#,+} & Teichmann, S.A.^{1,11,12,#,+}

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom.

2. University of Groningen, University Medical Center Groningen, Department of Pathology & Medical Biology, Groningen, The Netherlands. University of Groningen,

3. Groningen Research Institute for Asthma and COPD (GRIAC), University of Groningen, Groningen, The Netherlands.

4. University Medical Center Groningen, Department of Pulmonology, Groningen, The Netherlands

5. Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany.

6. Department of Surgery, University of Cambridge, and NIHR Cambridge Biomedical Research Centre, Cambridge, United Kingdom.

7. Allergic Inflammation Discovery Performance Unit, Respiratory Therapy Area, GlaxoSmithKline, Stevenage, United Kingdom.

8. Department of Mathematics, Technische Universität München, Munich, Germany.

9. Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Lung Biology and Disease, Group Systems Medicine of Chronic Lung Disease, and Translational Lung Research and CPC-M bioArchive, Member of the German Center for Lung Research (DZL), Munich, Germany.

10. University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, The Netherlands.

11. Open Targets, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom.

12. Theory of Condensed Matter Group, Cavendish Laboratory/Dept Physics, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0EH, UK

13. Department of Haematology, University of Cambridge, Cambridge, CB2 0XY, UK

14. Cambridge Stem Cell Institute, Cambridge CB2 1QR, UK

* These authors contributed equally to this work.

§ These authors contributed equally to this work.

These authors share senior authorship.

+ To whom correspondence should be addressed: m.c.nawijn@umcg.nl, st9@sanger.ac.uk

52 **Summary**

53 Human lungs enable efficient gas exchange, and form an interface with the environment
54 which depends on mucosal immunity for protection against infectious agents. Tightly
55 controlled interactions between structural and immune cells are required to maintain lung
56 homeostasis. Here, we use single cell transcriptomics to chart the cellular landscape of
57 upper and lower airways and lung parenchyma in health. We report location-dependent
58 airway epithelial cell states, and a novel subset of tissue-resident memory T cells. In lower
59 airways of asthma patients, mucous cell hyperplasia is shown to stem from a novel mucous
60 ciliated cell state, as well as goblet cell hyperplasia. We report presence of pathogenic
61 effector Th2 cells in asthma, and find evidence for type-2 cytokines in maintaining the altered
62 epithelial cell states. Unbiased analysis of cell-cell interactions identify a shift from airway
63 structural cell communication in health to a Th2-dominated interactome in asthma.

64

65

66

67 **Introduction**

68 The lung plays a critical role in both gas exchange and mucosal immunity, and its anatomy
69 serves these functions through (1) the airways that lead air to the respiratory unit, provide
70 mucociliary clearance, and form a barrier against inhaled particles and pathogens; and (2)
71 the alveoli, distal saccular structures where gas exchange occurs. Acute and chronic
72 disorders of the lung are a major cause of morbidity and mortality worldwide¹. To better
73 understand pathogenesis of lung disease, it is imperative to characterise the cell types of
74 the lung and understand their interactions in health^{2,3} and disease. The recent identification
75 of the ionocyte as a novel airway epithelial cell-type^{4,5} underscores our incomplete
76 understanding of the cellular landscape of the lung, which limits our insight into the
77 mechanisms of respiratory disease, and hence our ability to design therapies for most lung
78 disorders.

79
80 We set out to profile lung-resident structural and inflammatory cells and their interactions by
81 analysing healthy human respiratory tissue from four sources: nasal brushes, endobronchial
82 biopsies and brushes from living donors, and tissue samples from lung resections and
83 transplant donor lungs. Our single cell analysis identifies differences in the proportions and
84 transcriptional phenotype of structural and inflammatory cells between upper and lower
85 airways and lung parenchyma. Using an unbiased approach to identify tissue-resident CD4
86 T cells in airway wall, we identify a novel tissue migratory CD4 T cell (TMC) that harbours
87 features of both circulating memory cells and of tissue resident memory cells (TRM) CD4 T
88 cells. We demonstrate that many disease-associated genes have highly cell type-specific
89 expression patterns. This holds true for both rare disease-associated genes, such as CFTR
90 mutated in cystic fibrosis, as well as genes associated with a common disease such as
91 asthma.

92
93 In addition, we evaluate the altered cellular landscape of the airway wall in chronic
94 inflammatory disease using bronchial biopsies from asthma patients. We identify a novel
95 epithelial cell state highly enriched in asthma, the mucous ciliated cell. Mucous ciliated cells
96 represent a transitioning state of ciliated cells with molecular features of mucus production,
97 and contribute to mucous cell hyperplasia in this chronic disease. Other changes associated
98 with asthma include increased numbers of goblet cells, intraepithelial mast cells and
99 pathogenic effector Th2 cells in airway wall tissue. We examine intercellular communications
100 occurring in the healthy and asthmatic airway wall, and reveal a remarkable loss of epithelial
101 communication and a concomitant increase in Th2 cell interactions. The newly identified
102 TMC subset interacts with epithelial cells, fibroblasts and airway smooth muscle cells in
103 asthma. Collectively, these data generate novel insights into epithelial cell changes and
104 altered communication patterns between immune and structural cells of the airways, that
105 underlie asthmatic airway inflammation.

106 **A human lung cell census identifies macro-anatomical patterns of epithelial cell** 107 **states across the human the respiratory tree**

108 The cellular landscape along the 23 generations of the airways in human lung is expected
109 to differ both in terms of relative frequencies of cell types and their molecular phenotype⁶.
110 We used 10x Genomics Chromium droplet single-cell RNA sequencing (scRNA-Seq) to
111 profile a total of 36,931 single cells from upper and lower airways, and lung parenchyma
112 (Figure 1A, B). We profiled nasal brushes, and (bronchoscopic) brushes and biopsies from
113 airway wall (third to sixth generation) from healthy volunteers. For parenchyma (small
114 respiratory airways and alveoli), we obtained lung tissue from deceased transplant donors,
115 also analysed on the 10x platform, and from non-tumour resection tissue from lung cancer

116 patients, analysed on a bespoke droplet microfluidics platform based on the Dropseq
117 protocol⁷.

118
119 Integration of the data from nasal epithelium, airway wall and parenchymal tissue reveals a
120 diversity of epithelial, endothelial, stromal and immune cells, with approximately 21 coarse-
121 grained cell types in total (Figures 1 and 2, Extended Figure 1), that can be explored in user-
122 friendly web portal (www.lungcellatlas.org). Analysis of parenchymal lung tissue from
123 resection material using Dropseq led to the identification of 15 coarse-grained cell
124 populations (epithelial and non-epithelial) (Extended Figure 2). Using MatchScore⁸ to
125 quantify the overlap between cell type marker signatures between the two datasets revealed
126 an extensive degree of overlap in cell type identities (Extended Figure 2). In our analysis
127 below, we first concentrate on epithelial cells (Figure 1), and then focus on the stromal and
128 immune compartments (Figure 2).

129
130 In the epithelial lineage, we identified a total of at least 10 cell types across the upper and
131 lower airways and lung parenchyma (Figure 1C, Extended Data Figure 1). We detected
132 multiple basal, club, ciliated and goblet cell states, as well as type-1 (T1) and type-2 (T2)
133 alveolar cells, and the recently described ionocyte^{4,5} (Extended Figure 3). Both goblet and
134 ciliated cells were present in the nasal epithelium (Figure 1D). In the lower airways, we
135 detected basal, club and ciliated cells as well as ionocytes, but only very small numbers of
136 goblet cells. T1 and T2 cells were, as expected, only found in the lung parenchyma (Figure
137 1E).

138
139 We did not identify specific clusters of tuft cells or neuroendocrine (NE) cells. Since cell
140 types represented by a small fraction of the data might be missed by unsupervised
141 clustering, we evaluated the expression of known marker genes for NE cells (CHGA,
142 ASCL1, INSM1, HOXB5) and Tuft cells (DCLK1, ASCL2)⁴. NE marker genes identified a
143 small number of cells, present only in lower airways, displaying a transcriptional profile
144 consistent with that of NE cells (extended Figure 4). Tuft cell marker genes did not identify
145 a unique cell population. Ionocytes were found in lower airways, and at very low frequency
146 in upper airways, but were completely absent from the parenchyma. Comparison of the cell
147 populations identified using the two different bronchoscopic sampling methods (brush
148 versus biopsy) in lower airways showed that basal cells were captured most effectively in
149 biopsies, while apical epithelial cells, such as ciliated and club cells were relatively
150 overrepresented in the bronchial brushings (Figure 1D).

151
152 Our dataset allowed us to identify two discrete cell states in basal, goblet and ciliated
153 epithelial cells. Some of these cell phenotypes were restricted to specific anatomical
154 locations along the respiratory tract. Basal cells were present in both upper and lower
155 airways, although at relatively low frequency in upper airways (Figure 1E). The two basal
156 cell states corresponded to differentiation stages, with the less mature Basal 1 cell state
157 expressing higher levels of *TP63* and *NPPC* in comparison to Basal 2 cells (Figure 1F and
158 extended data 1), which were more abundant in bronchial brushes, suggesting a more apical
159 localization for these more differentiated basal cells (Figure 1D). Goblet 1 and 2 cells were
160 both characterized by high expression of *CEACAM5*, *S100A4*, *MUC5AC* and lack of *MUC5B*
161 (Figure 1F and Extended Figures 1 and 4). Goblet 1 cells specifically express *KRT4* and
162 *CD36* (Figure 1G and Extended Figure 4). Genes involved with immune function, such as
163 *IDO1*, *NOS2*, *IL19*, *CSF3* (Granulocyte-colony stimulating factor) and *CXCL10* are
164 expressed at high levels in Goblet 2 cells (Figure 1G and Extended Figure 4). These
165 molecules enriched in Goblet 2 cells are involved in recruitment of neutrophils, monocytes,
166 dendritic cells and T cells⁹. Both goblet cells states are present in upper airway epithelium,

167 with Goblet 1 cells being more frequent. In contrast, the Goblet 2 cell state was also present
168 in lower airways, albeit at low abundance (Figure 1E).

169
170 Ciliated cell transcriptional phenotypes are also zoned in terms of their presence across
171 macro-anatomical locations, with a discrete ciliated cell state more abundant in upper
172 airways (Ciliated 2) compared to lower airways and parenchyma. Nasal epithelial Ciliated 2
173 cells express pro-inflammatory genes, such as *CCL20* (Extended data 3) and higher levels
174 of metabolic genes (*ATP12A* and *COX7A1*) and vesicle transport (*AP2B1* and *SYT5¹⁰*)
175 compared to the Ciliated 1 cell state. In contrast, the Ciliated 1 cells from lower airways
176 specifically expressed genes involved in cytoprotection (*PROS1¹¹*) and fluid reabsorption
177 (*FXYD1¹²*) (Figure 1H and Extended Figure 4). Interestingly, comparison of the location-
178 specific differences between ciliated and goblet cells identified a transcriptional signature
179 specific for the upper airways present in both epithelial cell types (Extended Figure 4B).

180
181 Next, we assessed the contribution of specific epithelial cell types to Mendelian disease.
182 Cell-type specific expression patterns of genes associated with Mendelian disorders (based
183 on the Online Mendelian Inheritance in Man, OMIM database) confirm ionocytes as
184 particularly high expressers of the *CFTR* gene, mutated in cystic fibrosis (Figure 1I). These
185 cells also express *SCNN1B*, mutations of which can cause bronchiectasis, another feature
186 of cystic fibrosis, suggesting a potential key pathological role for ionocytes in both
187 bronchiectasis and cystic fibrosis. In addition, expression of *SERPINA1* (Figure 1I) was
188 found to be enriched in type-2 alveolar epithelial cells, underscoring their role in alpha-1-
189 antitrypsin deficiency¹³.

190 **Differential anatomical distribution of the stromal and immune components in the** 191 **human respiratory tree**

192 Next, we analysed the single cell transcriptomes of immune and stromal cells from the upper
193 airways, lower airways and the lung parenchyma (Figure 2A). We identified immune clusters
194 of myeloid (macrophages, neutrophils, dendritic cells (DCs) and mast cells) and lymphoid
195 cells (T and NK cells, B cells; Figure 2B, and Extended Figure 5). Immune and stromal cell
196 numbers and composition varied greatly across different anatomical regions (Figure 2A and
197 2C). Nasal brushes contained only a small number of immune cells, with the large majority
198 being dendritic cells. In the lower airways, the fraction of inflammatory cells was significantly
199 larger and relatively enriched for macrophages (Figure 2C and Extended Figure 5), which
200 was directly confirmed by cell composition comparison of upper *versus* lower airway brushes
201 obtained from the same donor (Extended Figure 5E).

202
203 Macrophages show large donor variation in their phenotype (Extended figure 5), but they all
204 share high expression of *MARCO*, *CCL18* and genes involved in apolipoprotein metabolism
205 (*APOC1* and *APOE*) (Figure 2E). Lung neutrophils express high levels of the granulocyte
206 markers *S100A8*, *S100A12¹⁴* and *LILRA5*, a receptor poorly characterised in the lungs, that
207 has been shown to have a proinflammatory function in synovial fluid macrophages¹⁵ (Figure
208 2E). DCs were mostly myeloid, with high expression of *CD1E*, *CD1C*, *CLEC10A* (Figure 2E)
209 and of *FCER1A* (IgE receptor) and *CCL17*, molecules known to play a key role in
210 inflammatory conditions such as asthma¹⁶.

211
212 In the droplet RNAseq data sets, we could not distinguish CD4+ and CD8+ T cells and NK
213 cells from each other (Figure 2B). The B cells in our dataset were mostly plasma cells,
214 expressing high levels of *JCHAIN* (Joining Chain of Multimeric IgA And IgM). IgM+ (*IGHM*)
215 cells were enriched in the airway lumen and in the lung parenchyma, while IgG3+ (*IGHG3*)
216 were enriched in airway biopsy samples and were virtually absent from the airway lumen.

217 This suggests an isotype-driven micro-anatomical segregation of B cells in the airways
218 (Extended Figure 5F).

219 **Molecular features of mucous cell metaplasia in asthma**

220 To characterize the changes in the cellular landscape of airway wall in a chronic
221 inflammatory condition, we also analysed bronchial biopsies from six volunteers with
222 persistent, childhood-onset asthma (Figure 3A). Asthma is a complex disease¹⁷ and multiple
223 cells such as epithelial^{18,19}, endothelial²⁰ and immune cells^{21,22} have been shown to be
224 altered in asthma. The combined airway wall dataset reveals a cellular landscape dominated
225 by epithelial (*EPCAM*-positive) cells, with minor contributions from endothelial,
226 mesenchymal and immune cells (Extended Figure 6A and B).

227
228 High-resolution clustering of the *EPCAM*⁺ clusters identifies 10 sub clusters representing the
229 6 epithelial cell types observed in healthy airway wall (Figure 1C), as well as two additional
230 basal cell states: a mucous ciliated cell state, and serous cells from the submucosal glands
231 (Figure 3B). In addition to the two basal cell states observed in healthy airway wall (Figure
232 1C), the basal cell states in asthma include activated and cycling cell states (Figure 3B).
233 Activated basal cells closely resemble Basal 1 cells in their transcriptional phenotype, but
234 also express proinflammatory genes such as *POSTN* (Figure 3D). Cycling basal cells are
235 characterized by expression of canonical marker genes of proliferating cells (*MKI67* and
236 *TOP2A*) (Figure 3D), and this is the only cluster of airway epithelial cells expressing the
237 squamous cell marker *KRT13* (Extended Figure 6).

238
239 We observe mucous cell hyperplasia in asthma, with a strong increase in goblet cell
240 numbers (Figure 3C), which are very rare in healthy airway wall biopsies (Figure 1E).
241 Moreover, the goblet cell transcriptional phenotype is altered in asthma, with strongly
242 increased expression of *MUC5AC* and *SPDEF*, as well as proinflammatory and remodelling
243 genes including *NOS2*, *CEACAM5* and *CST1* (Figure 3D). In addition, we identify a strong
244 increase in mucous ciliated cells, a novel cell state that has remarkable transcriptional
245 resemblance to ciliated cells, whilst co-expressing a number of mucous genes, including
246 *MUC5AC*, *SERPINB2/3* and *CEACAM5* (Figure 3D, Extended Figure 6). Mucous ciliated
247 cells lack expression of the transcription factor *SPDEF* (in contrast to club and goblet cells),
248 while maintaining *FOXJ1* expression, underscoring their ciliated cell origin (Extended Figure
249 7).

250
251 To further dissect the inferred differentiation trajectories in healthy and asthmatic airway wall
252 epithelial cells, we performed pseudotime analysis²³. This reveals a trajectory starting with
253 basal cell subsets, bifurcating into either a secretory lineage (mainly club cells) or a ciliated
254 lineage in healthy airway wall (Figure 3E). In asthma, the secretory lineage is a mix of club
255 and goblet cells, while the mucous ciliated cell state overlaps with the ciliated differentiation
256 trajectory (Figure 3E,F).

257
258 Next, we further analysed the transcriptional profiles of the two mucous cell states we
259 observe specifically in asthma: the mucous ciliated cells and the goblet cells. As both
260 NOTCH and IL4/IL13 signalling have been shown to contribute to mucous cell
261 differentiation²⁴, we analysed expression of both NOTCH target genes^{25,26} and IL4/IL13
262 target genes²⁷ in club, goblet, and ciliated cells as well as in the novel mucous ciliated cell
263 state, in both asthma and healthy airway wall biopsies. Expression of IL4/IL13-induced
264 genes²⁷ is prominent in asthma, and highest in activated basal cells, goblet cells and mucous
265 ciliated cells (Extended Figure 7). In club cells, expression of NOTCH target genes^{25,26} is
266 not different between asthma and healthy-derived cells. In contrast, in goblet cells, the

267 NOTCH target gene signature is retained only in cells from healthy airway wall, and is lost
268 in asthma.

269
270 As in goblet cells, mucous ciliated cells also lack expression of Notch target genes in asthma
271 (Extended Figure 7). Hence, we postulate that mucous ciliated cells represent a transition
272 cell state in the ciliated lineage - induced by IL4/IL13 signalling - leading to a mucous cell
273 phenotype which contributes to mucous cell metaplasia in asthma²⁴. Similar to goblet cells,
274 mucous ciliated cells express asthma genes such as *CST1*²⁸ and *POSTN* (Figure 3D),
275 indicating that these cells also contribute to airway inflammation and remodelling.

276
277 Integrating the asthma GWAS genes with our epithelial single cell transcriptomic data
278 reveals a broad contribution of the airway epithelial cell types to asthma susceptibility (Figure
279 3G), with high expression of asthma GWAS genes in ciliated and mucous ciliated cells. This
280 includes genes involved in cilia function (*DYNC2H1* and *KIF3A*), cell adhesion (*ELK3*,
281 *CDHR3* and *PTPRT*) and IL5-induced mucus metaplasia (*IL5RA*)²⁹, further suggesting a
282 direct link between mucous ciliated cells and Th2 CD4 T cells.

283 284 **Remodelling of the stromal and Immune compartments in asthmatic airways**

285 Asthma is associated with chronic inflammation and remodelling of the airway wall³⁰.
286 Analysis of the inflammatory and stromal cell populations in the bronchial biopsies by
287 unsupervised clustering (Figure 4A) reveals the presence of B and T cells, neutrophils,
288 macrophages, DCs, mast cells, fibroblasts, smooth muscle cells and endothelial cells
289 (Figure 4B, Extended Figure 8). We did not detect any innate lymphoid cells, basophils or
290 eosinophils as separate clusters (Extended Figure 8). Analysis of bulk transcriptome
291 analysis of whole airway biopsies before and after tissue dissociation identified very low
292 expression levels of eosinophil marker genes (*CLC* and *IL5RA*), indicating these cells are
293 relatively rare in the samples we analysed (Extended Figure 9).

294
295 Mast cell numbers were increased in asthma (Figure 4C), while being virtually absent in the
296 airways of healthy individuals (Figure 4C). Mast cells in asthmatic airways lack chymase 1
297 expression (*CMA1*) and express high levels of tryptase genes (*TPSB2*, *TPSAB1*) (Figure
298 4D). Prostaglandins and leukotrienes are known to be crucial to inflammatory cell signalling.
299 Mast cells express high levels of *PTGS2* and *HPGDS* (Figure 4D, Extended Figure 10).
300 *PTGS2* (cyclooxygenase-2), also known as inflammatory cyclooxygenase, converts the
301 precursor arachidonic acid to prostaglandin endoperoxide H₂ (PGH₂). *HPGDS*
302 (Hematopoietic Prostaglandin D Synthase) catalyses the conversion of PGH₂ to
303 prostaglandin D₂ (PGD₂). PGD₂ activates CD4 Th2 cells³¹, ILC2³², basophils and
304 neutrophils³¹ and plays a key role in asthma pathology. Expression of all PGD₂ biosynthesis
305 enzymes is a unique feature of mast cells (Extended Figure 10) and this suggests that
306 intraepithelial mast cells are continually producing PGD₂ in asthma patients. Thus, these
307 cells are most likely intraepithelial mast cells, previously shown to accumulate in Th2-high
308 asthmatic airway epithelium³³, and reported to be increased³⁴ with disease severity²¹.

309
310 We observed an increase in the number of B cells in the asthmatic airways (Figure 4C) and
311 these cells have a plasma cell phenotype, with high *JCHAIN* expression (Figure 4D). The
312 increase in B cell numbers was mostly of IgM+ cells (*IGHM*) (Figure 4E). IgM levels in
313 asthma BALF samples have been reported to be either increased³⁴ or unchanged³⁵,
314 suggesting cohort dependent variability. IgM-producing B cells in the healthy airways were
315 mostly present in the airway lumen (Extended Figure 5F). However, as we did not analyse
316 brush samples from asthmatic patients, we cannot precisely pinpoint whether the increase
317 in IgM+ B cells takes place in the intraepithelial region or in the lumen, as both regions are
318 present in biopsy samples.

319

320 Asthma GWAS genes show highly cell-type restricted expression (Figure 4F). When
321 excluding the widely expressed HLA genes from the analysis, fibroblasts and T cells express
322 the highest number of asthma GWAS genes (Figure 4F), which are mostly upregulated in
323 asthma (Figure 4F). GATA3 expression is restricted to T cells (Figure 4F), and increased in
324 T cells from asthma patients (Figure 4F and G). We detected increased expression of CD4
325 (but not CD8a) in the T cell cluster, suggesting an increase in Th2 CD4 T cells (Figure 4G).
326 Therefore, we next proceed to investigate the CD4 T cell compartment in depth.

327 **Pathogenic effector Th2 cells are enriched in asthmatic airways**

328 In line with the increase in GATA3 and CD4 expression mentioned above, CD4 Th2 cells
329 are known to be key drivers of asthma^{17,36}. To assess the presence of Th2 effector cells in
330 the airways of asthma patients (Figure 4G), we single cell sorted CD4 T cells followed by
331 plate-based SmartSeq2 analysis for in depth transcriptional phenotyping of the T helper cell
332 compartment (see Methods for details). We analysed cells from both peripheral blood and
333 airway wall biopsies (Figure 5A) from a larger cohort of asthma patients and healthy controls
334 (Figure 5B). Unbiased clustering reveals six major populations of CD4 T cells (Figure 5C
335 and Extended Figure 11). At this coarse level of analysis, none of these six clusters was
336 specifically enriched in asthma patients (Figure 5D).

337
338 Comparative analysis of CD4 T cells isolated from paired blood and lung samples allows us
339 to differentiate between tissue-resident T cells and circulating T cells in an unbiased way,
340 by subtracting the populations shared with blood from the populations specific to the
341 biopsies (Figure 5E). Using this approach, we identified two subsets highly enriched in the
342 lungs: the classical Tissue Resident Memory (TRM) CD4 T cells, and a novel subset we
343 named Tissue Migratory CD4 T cell (TMC) (Figure 5E). Naive/central memory (CM), effector
344 memory (EM), and EMRA cells, as well as a mixed Treg/Th2 cluster, are either enriched in
345 blood or present in both blood and airways (Figure 5E).

346
347 To better understand the two distinct lung-restricted CD4 T cell subsets, we performed
348 differential expression analysis between TRM and TMC cells (Figure 5F). Several
349 transcription factors highly expressed in circulating cells are enriched in TMC cells, such as
350 *LEF1*, *SATB1* and *KLF3*, while *ZEB2* is specific for TRM cells. TMC cells expressed the
351 tissue egression markers *S1PR1*, *CCR7* and *SELL* (CD62L) and lacked expression of the
352 canonical TRM marker *ITGAE* (CD103) (Figure 5F). As low numbers of TMC cells were
353 present in peripheral blood CD4 T cells (Figure 5E), these data suggest that these cells have
354 the potential to transit between lung and blood, a hypothesis supported by pseudotime
355 trajectory analysis of the CD4 T cell subsets (Extended Figure 12).

356
357 Protein expression of CD69 and CD103 (ITGAE) have both been used as hallmarks of lung
358 resident CD4 T cells isolated from lung parenchyma^{37,38}, but to the best of our knowledge,
359 no similar analysis of lung airway epithelial CD4 T cells has been performed to date. Both
360 TMC and TRM cells have high expression of CD69, but only TRM cells express ITGAE
361 (Extended Figure 11), suggesting these subsets might be equivalent to the previously
362 described cells^{37,38}. However, TRM cells^{37,38} have been shown to lack S1PR1 and CCR7
363 protein expression, in contrast to TMC cells that express high mRNA levels of both markers.
364 No direct whole transcriptome comparison of CD69+CD103+ *versus* CD69+CD103- TRMs
365 has been performed yet^{37,38}. Further studies are necessary to properly compare CD4 T cells
366 from airway wall *versus* lung parenchyma, and investigate how TMC align with the previously
367 reported TRM subsets at the transcriptome level.

368

369 TRM cells in airway wall expressed high levels of *CXCR6* and *ITGA1* and high levels of
370 cytokines (*CCL4*, *CCL4L2*, *CCL5*) and effector molecules (*PRF1*, *GZMB*, *GZMA*, *GZMH*)
371 (Figure 5F and Extended Figure 11), indicating they are also in a primed state capable of
372 direct effector function, as recently shown for TRMs from lung parenchyma³⁷.

373
374 CD4 effector T cells are classically divided into distinct functional subsets based on their
375 cytokine profile¹⁷. We manually annotated clusters of Th1 (*IFNG*⁺), Th2 (*IL4*⁺, *IL5*⁺ or *IL13*⁺)
376 and Th17 (*IL17A*⁺ or *IL17F*⁺) cells based on their cytokine expression profiles (Figure 5G,
377 Extended Figure 13). Cytokine-producing cells were mostly retrieved from the biopsies, with
378 only very few present in blood (Extended Figure 14).

379
380 In terms of absolute numbers, Th2 cells were very rare and found both in healthy and
381 asthmatic patients, although numbers of Th2 cells were significantly increased in the airway
382 wall in the asthma patients, with no detected difference in the relative proportions of the
383 other subsets (Figure 5H). In addition to the signature cytokines and the transcription factor
384 *GATA3* (Extended Figure 13), airway wall Th2 cells express *HGPDS*, identifying them as
385 pathogenic effector Th2 (peTh2) cells, previously associated with eosinophilic inflammation
386 of the gastrointestinal tract and skin³⁹. Airway Th2 also express the transcription factor
387 *PPARG*, and cytokine receptors *IL17RB* and *IL1RL1* (Figure 5I). *IL17RB* and *IL1RL1* have
388 been reported as upregulated in pathogenic allergen-specific Th2 cells (coined Th2A cells),
389 which are present in allergic disease⁴⁰ as well as in chronic rhinosinusitis with nasal polyps⁴¹,
390 suggesting airway wall Th2 cells share features with both Th2A and peTh2 cells.

391 **Asthma is characterized by specific signalling networks**

392 Asthma is characterized by remodelling of the airways, which depends on complex
393 interactions between structural and inflammatory cells¹⁷, both *via* direct physical interactions
394 and *via* secreted proteins and small molecules. We used our recently developed
395 receptor/ligand database and statistical inference framework CellPhoneDB⁴²
396 (www.cellphonedb.org), to chart combinatorial cell-specific expression patterns of genes
397 encoding receptor/ligand pairs. We aim to identify potential cell-cell interactions in the airway
398 wall, and define their changes in asthma. Whilst most interactions are unchanged, some
399 were specific to the diseased or healthy state (Full list in Extended 6).

400
401 In healthy controls, the cell-cell interaction landscape of the airway wall was dominated by
402 lung structural cells (mainly mesenchymal and epithelial cell types) communicating with
403 other lung structural cells and with tissue-resident T cells, both the classical TRM and the
404 newly identified TMC subsets (Figure 6A,B, left panels). In the asthmatic airway wall, the
405 number of predicted interactions between epithelial and mesenchymal cells was strongly
406 reduced. Instead, the cell-cell communication landscape in asthmatic airway wall is
407 dominated by Th2 cells that were found to have increased interactions with other immune
408 cells, including antigen-presenting cells, and also with epithelial cells ((Figure 6A,B, right
409 panels). The most striking increase in interactions is with mesenchymal cells, both
410 fibroblasts and smooth muscle cells (Figure 6A and B, right panels).

411
412 Analysis of the predicted cell-cell interactions between structural cells in healthy airway wall
413 revealed a wealth of growth factor-signalling pathways including the FGF, GFR, IGF, TGF,
414 PDGF and VEGF pathways, most of which were lost in the asthmatic samples (Figure 6C).
415 Detailed analysis of the individual interactions of Th2 cells with the inflammatory and lung
416 structural cells unique to asthma revealed the potential of Th2 cells to have cognate
417 interactions with the epithelial cells involving KLRG1 and CD103 binding to E-cadherin, and
418 also integrin-mediated interactions with epithelial-expressed matrix proteins such as

419 Tenascin-C. Epithelial expression of alarmins and cytokines, such as IL33, TSLP (Figure
420 3G) and TNFSF10/TRAIL (Figure 6D), all of which are asthma genes⁴³⁻⁴⁵, might then lead
421 to activation of Th2 cells expressing the receptors.
422

423 In addition to validating these well-known interactions, which for IL33 and TSLP failed to
424 reach significance in our unbiased cell-cell interaction analysis, we identify novel epithelial-
425 Th2 cell interactions in asthma, including chemokines CXCL2 and CXCL17, and the cytokine
426 MIF that are all expressed by epithelial cells, while Th2 cells express the respective
427 receptors (Figure 6C). Interestingly, mesenchymal cells share some of these Th2 cell
428 interactions, such as expression of TNFSF10/TRAIL and MIF. Predicted Th2 interactions
429 unique to mesenchymal cells in asthmatic airway wall are CXCL12 and CCL11, expressed
430 by fibroblasts and smooth muscle cells. Airway wall Th2 cells in asthma express the
431 cytokines IL5 and IL13 (Figure 5I), the receptor complexes for which are expressed by
432 immune cells and epithelial cells, respectively, in line with the expression of IL13-driven
433 genes in mucous ciliated and goblet cells in asthma (Extended data Figure 7). In addition to
434 these classical Th2 cytokines, Th2 cells express LTB for which basal epithelial cells express
435 the receptor.
436

437
438
439
440
441

442 Discussion

443

444 Our study profiles the cellular landscape of human lung tissue at the single-cell level,
445 including both upper and lower airways and parenchymal lung tissue in healthy adults. We
446 identify at least 21 main cell types in the normal human lung, that can be further subdivided
447 into more fine-grained cell states. There is clear Mendelian disease relevance for many cells,
448 including the previously-reported ionocytes (for bronchiectasis and cystic fibrosis), and type-
449 2 alveolar cells (for alpha-1-antitrypsin deficiency) . We chart differences in frequencies and
450 molecular state of airway epithelial cells between upper and lower airways. To our
451 knowledge, we provide for the first time a detailed molecular description of tissue-resident
452 CD4 T cells in the human lower airway wall, and identify two separate subsets, one of which
453 was hitherto unknown.

454

455 In our studies, we deployed two different droplet-based single-cell RNA sequencing
456 platforms and one plate-based method, with experiments performed in three geographically
457 distinct research centers. Overall, the datasets are remarkably consistent, which yields a
458 quantitative dataset of the cellular composition of human upper and lower airways and lung
459 parenchyma.

460

461 In addition to analysing healthy reference samples, we characterize the changes in the cell
462 types and cell states in airway wall in asthma. This reveals mucous ciliated cells as a novel
463 cell state that contributes to mucous cell metaplasia. Both mucous ciliated and goblet cells
464 are characterized by expression of an IL4/IL13-driven gene signature, indicating a dominant
465 role for type-2 cytokines in maintaining the epithelial changes in chronic airway inflammation
466 in asthma. The mucin gene expression induced in FOXP1+ cells with an unabated ciliated
467 cell transcriptional profile strongly indicates that the mucous state is superimposed on the
468 ciliated cell phenotype, independent of goblet cell differentiation from club cells. Our data
469 seem to indicate that these two processes can occur in parallel, with mucous metaplasia of
470 ciliated cells and goblet cell hyperplasia both contributing to the increase in mucin-producing
471 cells in asthma. Whether the mucous ciliated cells go on to lose their FOXP1 expression and
472 the ciliated transcriptional profile and transdifferentiate into *bona fide* SPDEF-positive goblet
473 cells remains to be firmly established.

474

475 The epithelial cell changes in airway wall in asthma are surprisingly different to those
476 recently described in patients with in chronic rhinosinusitis with polyps²⁷. In this chronic
477 inflammatory disease of the upper airways caused by exaggerated type-2 immunity, an
478 IL4/IL13-driven gene transcription profile was mainly observed in basal epithelial cells, which
479 were found to be arrested in differentiation and highly increased in numbers²⁷. Using the
480 same IL4/IL13-driven gene module, we find some expression thereof in basal cells, but in
481 asthma this does not result in significant changes in basal cell numbers. Instead, we observe
482 an increased number of goblet cells, as well as of mucous ciliated cells, both of which show
483 evidence of marked expression of the IL4/IL13-driven gene signature (extended data Figure
484 7). Hence, while there is some overlap in the cellular mechanisms underlying rhinosinusitis
485 with polyps and asthma, the resultant changes in cellular states and their frequencies in
486 airway wall differ considerably between the epithelia of the upper *versus* the lower airways.
487 In contrast, the changes in the eicosanoid pathway observed in chronic type-2 inflammation
488 of the upper²⁷ and lower (extended data Figure 10) airways are very similar, likely reflecting
489 a common cellular mechanism between Th2 inflammation in these two anatomical locations.

490

491 Conflicting data have been reported on dependence of IL13-induced goblet cell metaplasia
492 on NOTCH signalling in *ex vivo* cultured PBECs^{25,46}. Our data on freshly isolated bronchial

493 epithelial cells show that expression of NOTCH2 and Notch target genes is present in goblet
494 cells from healthy controls only, and absent from ciliated, mucous ciliated and goblet cells
495 in asthmatic airway wall. This indicates that mucous cell metaplasia in mild asthmatics is
496 likely to be driven by type-2 cytokines in a NOTCH-independent fashion. One likely source
497 of the type-2 cytokines driving this goblet cell metaplasia are the cytokine-expressing
498 effector Th2 cells that are increased in asthmatic airway wall. To our knowledge, our study
499 is the first to conclusively show the presence of the recently identified^{39,40,41} pathogenic
500 effector Th2 cells in the airway wall in asthma, as evidenced by the combined expression of
501 *IL5*, *IL13*, *HPGDS*, *PPARG*, *IL17RB*, *IL1RL1* and *IL1RAP*. Additional cellular sources for the
502 type-2 cytokines, including innate lymphoid cells, cannot be ruled out based on our data, as
503 these cells were present in too low numbers to be analysed in our dataset, and will need to
504 be purified from the biopsy cell suspensions for further characterization in future studies.
505

506 Finally, our detailed and unbiased analyses of cell-cell communication in healthy and
507 asthmatic airway wall reveals novel interactions of the airway wall resident cells in health
508 and disease. Comprehensive analysis of the cell-cell interactions underpinning the changes
509 of the airway wall cellular landscape in asthma identifies a shift away from interactions
510 between structural cells in healthy airway wall, towards an intercellular network dominated
511 by the interactions of Th2 cells with structural and inflammatory cells in asthmatic airway
512 wall. The richness of growth factor signalling between epithelial cells and mesenchymal cells
513 observed in healthy airway wall is largely lost in asthma, which seems at odds with a
514 reactivation of the epithelial-mesenchymal trophic unit in asthma⁴⁷. Instead, our data
515 supports a shift in cellular phenotypes in airway wall due to the local production of Th2
516 cytokines such as IL13 in chronic disease in our patient cohort with childhood-onset asthma.
517 This global view of the airway wall cellular landscape in health and in asthma opens up new
518 perspectives on lung biology and molecular mechanisms of asthma.
519

520 **Acknowledgments:**

521 We thank Jana Elias (scientific illustrator) for the design of cartoons, the Sanger Single Cell
522 Genomics Core Facility for support with the SmartSeq2 protocol, Dr. Emma Rawlins for
523 feedback and critical reading of the manuscript, as well as all the members of the Teichmann
524 lab for scientific input. We are grateful to the Cambridge Biorepository for Translational
525 Medicine for the provision of tissue from deceased organ donors.

526
527 **Authors contribution:**

528
529 Designed the project: Teichmann, S.A, Nawijn, M.C., van den Berge, M., Affleck, K., van
530 Oosterhout, A.J., Schiller, H.B. Wrote the paper: Vieira Braga, F.A., Nawijn, M.C.,
531 Teichmann, S.A. Generated data: Vieira Braga, F.A., Carpaij, O.A., Brouwer, S., Hesse, L.,
532 Jiang, J., Fasouli, E.S., Strzelecka, P., Mahbubani, K.T.A., Angelidis, I., Strunz, M. Analyzed
533 data: Vieira Braga, F.A., Kar, G., Berg, M. Simon, L., Gomes, T., Jiang, J., Efremova, M.,
534 Palit, S., Polanski, K., Firth, H.V., Theis, F.J. Interpreted data: Vieira Braga, F.A., Kar, G.,
535 Carpaij, O.A., Simon, L., Gomes, T., Jiang, J., Vento-Tormo, R., Affleck, K., Palit, S., Cvejic,
536 A., Saeb-Parsy, K., Timens, W., Koppelman, G.H., van Oosterhout, A.J., Schiller, H.B., van
537 den Berge, M., Theis, F.J., van den Berge, M., Meyer, K.B. All authors read the manuscript,
538 offered feedback and approved it prior to submission.

539
540 **Funding:**

541 This work was funded by Open Targets, an open innovation public-private partnership
542 (<http://www.opentargets.org>), a GlaxoSmithKline collaborative agreement with University
543 Medical Center Groningen, Wellcome (WT206194), EMBO and HFSP Long Term
544 fellowships to R. Vento-Tormo, the Marie Curie ENLIGHT-TEN training network for Tomas
545 Gomes, the Lung Foundation Netherlands (projects no 5.1.14.020 and 4.1.18.226), and
546 Health-Holland, Top Sector Life Sciences & Health. LMS acknowledges funding from the
547 European Union's Horizon 2020 research and innovation programme under the Marie
548 Sklodowska-Curie grant agreement No 753039.

549
550 **Competing interests.**

551 Affleck, K and van Oosterhout, A.J are employees of GSK.

552
553 **Data availability:**

554 Interactive exploration tool: www.lungcellatlas.org

555
556 **Materials & Correspondence.**

557 m.c.nawijn@umcg.nl, st9@sanger.ac.uk

558

559

560 References

- 561 1. Bousquet, J., Dahl, R. & Khaltaev, N. Global Alliance against Chronic Respiratory Diseases.
562 *Eur. Respir. J.* **29**, 233–239 (2007).
- 563 2. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, (2017).
- 564 3. Franks, T. J. *et al.* Resident cellular components of the human lung: current knowledge and
565 goals for research on cell phenotyping and function. *Proc. Am. Thorac. Soc.* **5**, 763–766
566 (2008).
- 567 4. Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing
568 ionocytes. *Nature* **560**, 319–324 (2018).
- 569 5. Plasschaert, L. W. *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich
570 pulmonary ionocyte. *Nature* **560**, 377–381 (2018).
- 571 6. Tata, P. R. & Rajagopal, J. Plasticity in the lung: making and breaking cell identity.
572 *Development* **144**, 755–766 (2017).
- 573 7. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells
574 Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
- 575 8. Mereu, E. *et al.* matchScore: Matching Single-Cell Phenotypes Across Tools and
576 Experiments. *bioRxiv* 314831 (2018). doi:10.1101/314831
- 577 9. Bisset, L. R. & Schmid-Grendelmeier, P. Chemokines and their receptors in the pathogenesis
578 of allergic asthma: progress and perspective. *Curr. Opin. Pulm. Med.* **11**, 35–42 (2005).
- 579 10. Colvin, R. A. *et al.* Synaptotagmin-mediated vesicle fusion regulates cell migration. *Nat.*
580 *Immunol.* **11**, 495–502 (2010).
- 581 11. Urawa, M. *et al.* Protein S is protective in pulmonary fibrosis. *J. Thromb. Haemost.* **14**, 1588–
582 1599 (2016).
- 583 12. Wujak, Ł. A. *et al.* FXYP1 negatively regulates Na(+)/K(+)-ATPase activity in lung alveolar
584 epithelial cells. *Respir. Physiol. Neurobiol.* **220**, 54–61 (2016).
- 585 13. Krotova, K. *et al.* Alpha-1 Antitrypsin-Deficient Macrophages Have Increased Matriptase-
586 Mediated Proteolytic Activity. *Am. J. Respir. Cell Mol. Biol.* **57**, 238–247 (2017).
- 587 14. Vogl, T. *et al.* S100A12 is expressed exclusively by granulocytes and acts independently from
588 MRP8 and MRP14. *J. Biol. Chem.* **274**, 25291–25296 (1999).
- 589 15. Mitchell, A. *et al.* LILRA5 is expressed by synovial tissue macrophages in rheumatoid arthritis,
590 selectively induces pro-inflammatory cytokines and IL-10 and is regulated by TNF-alpha, IL-10
591 and IFN-gamma. *Eur. J. Immunol.* **38**, 3459–3473 (2008).
- 592 16. Condon, T. V., Sawyer, R. T., Fenton, M. J. & Riches, D. W. H. Lung dendritic cells at the
593 innate-adaptive immune interface. *J. Leukoc. Biol.* **90**, 883–895 (2011).
- 594 17. Holgate, S. T. *et al.* Asthma. *Nat Rev Dis Primers* **1**, 15025 (2015).
- 595 18. Lopez-Guisa, J. M. *et al.* Airway epithelial cells from asthmatic children differentially express
596 proremodeling factors. *J. Allergy Clin. Immunol.* **129**, 990–7.e6 (2012).
- 597 19. Alcalá, S. E. *et al.* Mitotic asynchrony induces transforming growth factor-β1 secretion from
598 airway epithelium. *Am. J. Respir. Cell Mol. Biol.* **51**, 363–369 (2014).
- 599 20. Harkness, L. M., Ashton, A. W. & Burgess, J. K. Asthma is not only an airway disease, but
600 also a vascular disease. *Pharmacol. Ther.* **148**, 17–33 (2015).
- 601 21. Balzar, S. *et al.* Mast cell phenotype, location, and activation in severe asthma. Data from the
602 Severe Asthma Research Program. *Am. J. Respir. Crit. Care Med.* **183**, 299–309 (2011).
- 603 22. Truyen, E. *et al.* Evaluation of airway inflammation by quantitative Th1/Th2 cytokine mRNA
604 measurement in sputum of asthma patients. *Thorax* **61**, 202–208 (2006).
- 605 23. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by
606 pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
- 607 24. Erle, D. J. & Sheppard, D. The cell biology of asthma. *J. Cell Biol.* **205**, 621–631 (2014).
- 608 25. Danahay, H. *et al.* Notch2 is required for inflammatory cytokine-driven goblet cell metaplasia in
609 the lung. *Cell Rep.* **10**, 239–252 (2015).
- 610 26. Gomi, K., Arbelaez, V., Crystal, R. G. & Walters, M. S. Activation of NOTCH1 or NOTCH3
611 signalling skews human airway basal cell differentiation toward a secretory pathway. *PLoS*
612 *One* **10**, e0116507 (2015).
- 613 27. Ordovas-Montanes, J. *et al.* Allergic inflammatory memory in human respiratory epithelial
614 progenitor cells. *Nature* **560**, 649–654 (2018).
- 615 28. Luo, W. *et al.* Airway Epithelial Expression Quantitative Trait Loci Reveal Genes Underlying

- 616 Asthma and Other Airway Diseases. *Am. J. Respir. Cell Mol. Biol.* **54**, 177–187 (2016).
- 617 29. Wu, C. A. *et al.* Bronchial epithelial cells produce IL-5: implications for local immune
618 responses in the airways. *Cell. Immunol.* **264**, 32–41 (2010).
- 619 30. Laitinen, L. A., Laitinen, A. & Haahtela, T. Airway mucosal inflammation even in patients with
620 newly diagnosed asthma. *Am. Rev. Respir. Dis.* **147**, 697–704 (1993).
- 621 31. Arima, M. & Fukuda, T. Prostaglandin D2 and TH2 Inflammation in the Pathogenesis of
622 Bronchial Asthma. *Korean J. Intern. Med.* **26**, 8 (2011).
- 623 32. Xue, L. *et al.* Prostaglandin D2 activates group 2 innate lymphoid cells through
624 chemoattractant receptor-homologous molecule expressed on TH2 cells. *J. Allergy Clin.*
625 *Immunol.* **133**, 1184–1194 (2014).
- 626 33. Dougherty, R. H. *et al.* Accumulation of intraepithelial mast cells with a unique protease
627 phenotype in T(H)2-high asthma. *J. Allergy Clin. Immunol.* **125**, 1046–1053.e8 (2010).
- 628 34. Hol, B. E., van de Graaf, E. A., Out, T. A., Hische, E. A. & Jansen, H. M. IgM in the airways of
629 asthma patients. *Int. Arch. Allergy Appl. Immunol.* **96**, 12–18 (1991).
- 630 35. Peebles, R. S., Jr, Liu, M. C., Lichtenstein, L. M. & Hamilton, R. G. IgA, IgG and IgM
631 quantification in bronchoalveolar lavage fluids from allergic rhinitics, allergic asthmatics, and
632 normal subjects by monoclonal antibody-based immunoenzymetric assays. *J. Immunol.*
633 *Methods* **179**, 77–86 (1995).
- 634 36. Muehling, L. M., Lawrence, M. G. & Woodfolk, J. A. Pathogenic CD4+ T cells in patients with
635 asthma. *J. Allergy Clin. Immunol.* **140**, 1523–1540 (2017).
- 636 37. Oja, A. E. *et al.* Trigger-happy resident memory CD4+ T cells inhabit the human lungs.
637 *Mucosal Immunol.* **11**, 654–667 (2018).
- 638 38. Kumar, B. V. *et al.* Human Tissue-Resident Memory T Cells Are Defined by Core
639 Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep.* **20**,
640 2921–2934 (2017).
- 641 39. Mitson-Salazar, A. *et al.* Hematopoietic prostaglandin D synthase defines a proeosinophilic
642 pathogenic effector human T(H)2 cell subpopulation with enhanced function. *J. Allergy Clin.*
643 *Immunol.* **137**, 907–18.e9 (2016).
- 644 40. Wambre, E. *et al.* A phenotypically and functionally distinct human TH2 cell subpopulation is
645 associated with allergic disorders. *Sci. Transl. Med.* **9**, (2017).
- 646 41. Lam, E. P. S. *et al.* IL-25/IL-33-responsive TH2 cells characterize nasal polyps with a default
647 TH17 signature in nasal mucosa. *J. Allergy Clin. Immunol.* **137**, 1514–1524 (2016).
- 648 42. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal–fetal interface in
649 humans. *Nature* **563**, 347–353 (2018).
- 650 43. Weckmann, M., Kopp, M. V., Heinzmann, A. & Mattes, J. Haplotypes covering the TNFSF10
651 gene are associated with bronchial asthma. *Pediatr. Allergy Immunol.* **22**, 25–30 (2011).
- 652 44. Harada, M. *et al.* Thymic stromal lymphopoietin gene promoter polymorphisms are associated
653 with susceptibility to bronchial asthma. *Am. J. Respir. Cell Mol. Biol.* **44**, 787–793 (2011).
- 654 45. Grotenboer, N. S., Ketelaar, M. E., Koppelman, G. H. & Nawijn, M. C. Decoding asthma:
655 translating genetic variation in IL33 and IL1RL1 into disease pathophysiology. *J. Allergy Clin.*
656 *Immunol.* **131**, 856–865 (2013).
- 657 46. Gerovac, B. J. & Fregien, N. L. IL-13 Inhibits Multicilin Expression and Ciliogenesis via Janus
658 Kinase/Signal Transducer and Activator of Transcription Independently of Notch Cleavage.
659 *Am. J. Respir. Cell Mol. Biol.* **54**, 554–561 (2016).
- 660 47. Holgate, S. T. *et al.* Epithelial-mesenchymal communication in the pathogenesis of chronic
661 asthma. *Proc. Am. Thorac. Soc.* **1**, 93–98 (2004).
- 662 48. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell
663 transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**,
664 411–420 (2018).
- 665 49. Song, J. *et al.* Aberrant DNA methylation and expression of SPDEF and FOXA2 in airway
666 epithelium of patients with COPD. *Clin. Epigenetics* **9**, 42 (2017).
- 667 50. Heijink, I. H. *et al.* Down-regulation of E-cadherin in human bronchial epithelial cells leads to
668 epidermal growth factor receptor-dependent Th2 cell-promoting activity. *J. Immunol.* **178**,
669 7678–7685 (2007).
- 670 51. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–
671 181 (2014).
- 672 52. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short

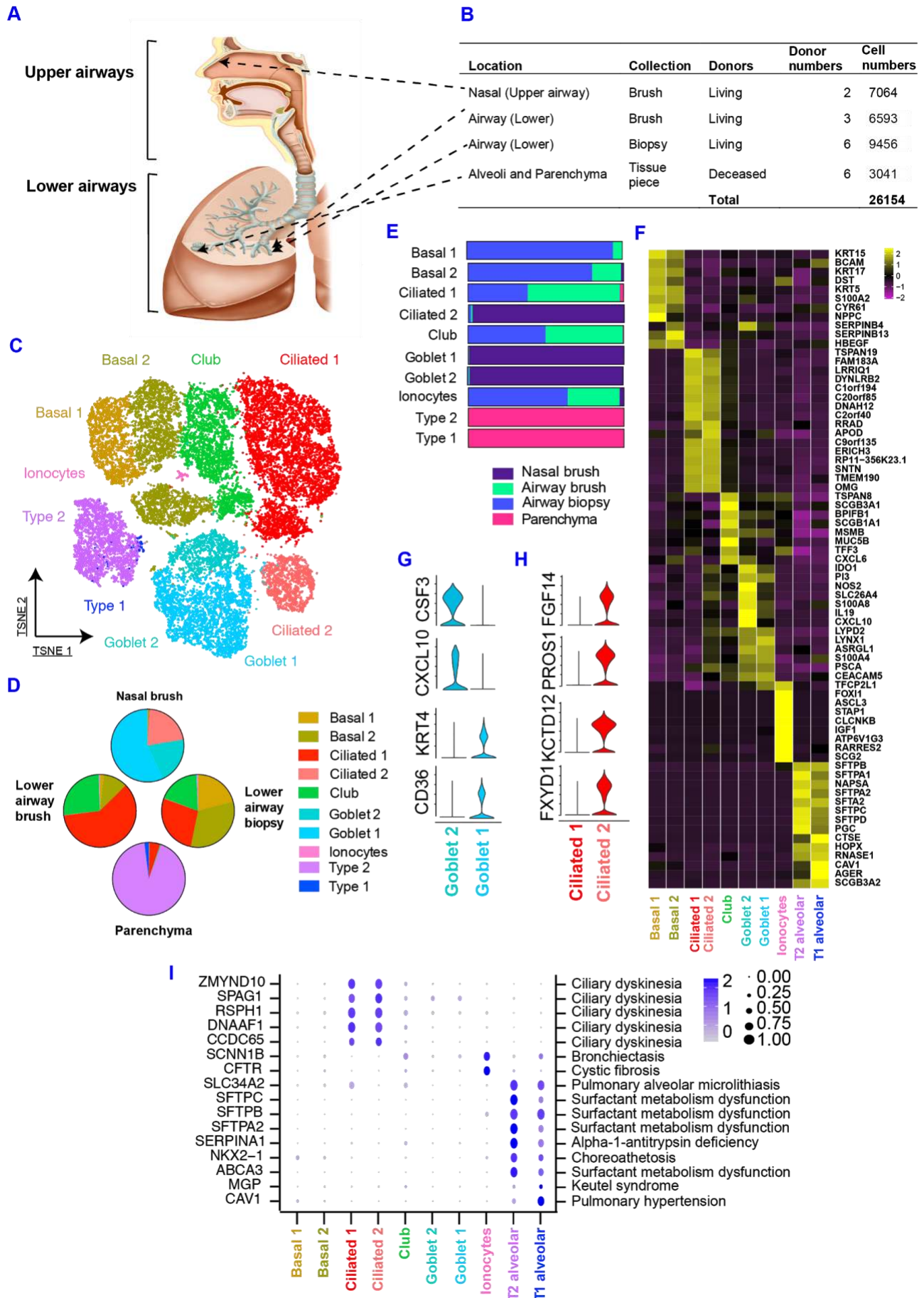
- 673 reads. *Bioinformatics* **26**, 873–881 (2010).
- 674 53. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- 675 54. van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene
- 676 expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).
- 677 55. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet based
- 678 single cell RNA sequencing data. *bioRxiv* (2018).

679

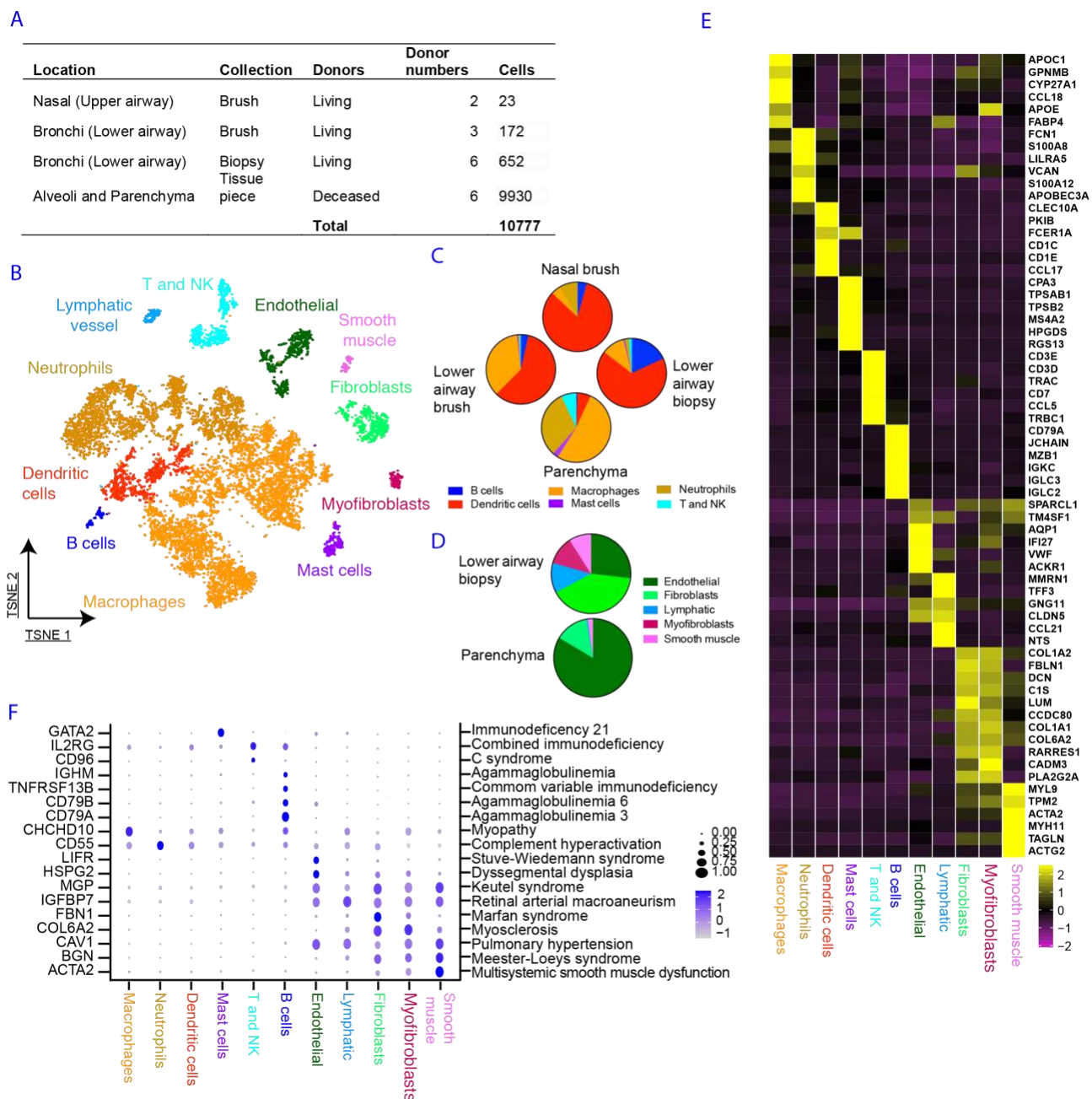
680

681

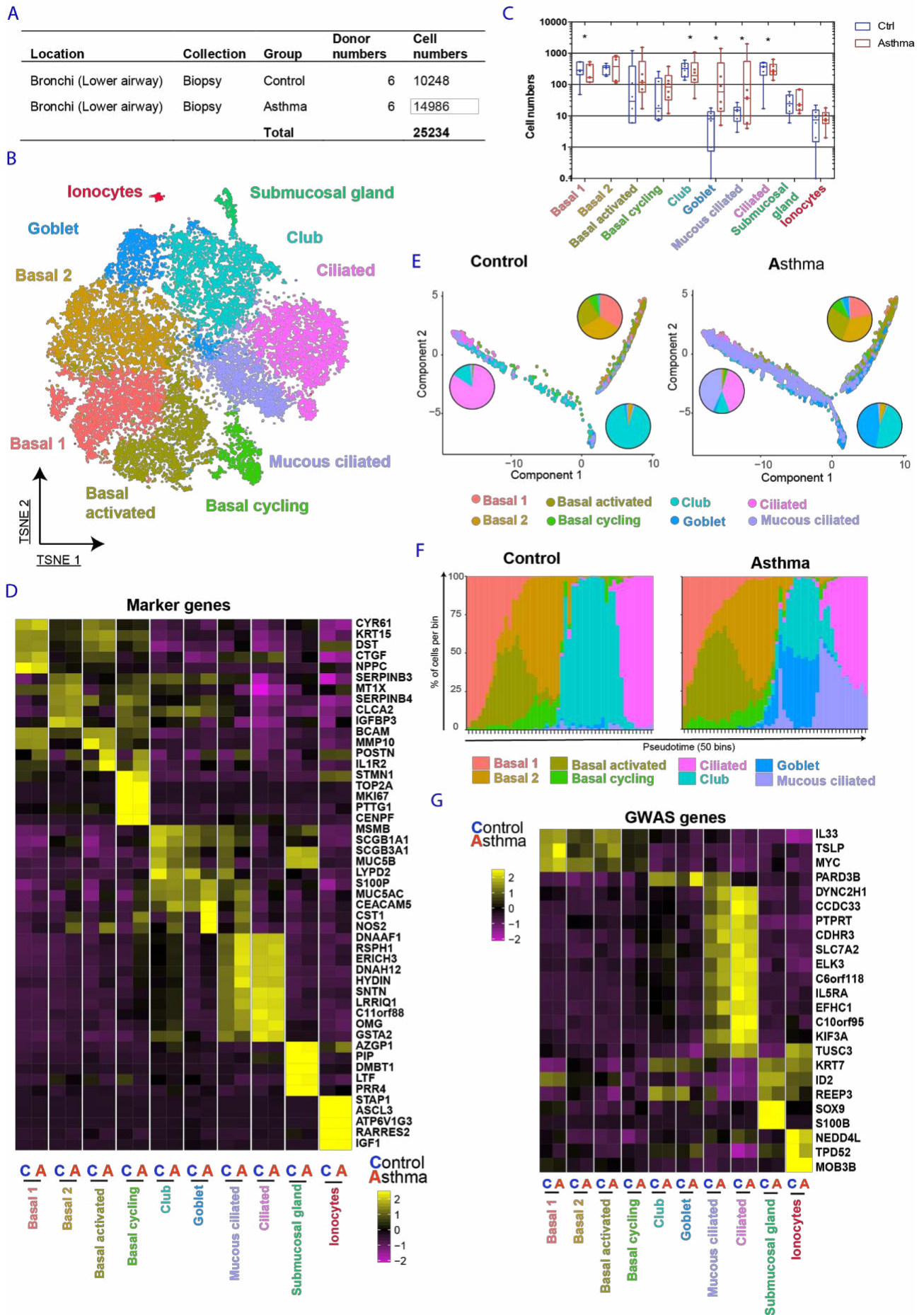
682



684 **Figure 1. A human lung cell census identifies zonation of novel epithelial cell states**
685 **across macro-anatomical location. (A)** Schematic illustration depicting anatomical
686 regions analysed in this manuscript. **(B)** Table with the details of anatomical region, tissue
687 source, donors and cell numbers present in this figure. **(C)** tSNE displaying the major
688 epithelial clusters present in the full extent of the human respiratory tree. **(D)** Pie charts
689 depicting the cellular composition by anatomical region. **(E)** Horizontal slice bar depicting
690 the anatomical distribution of each cell type identified **(F)** Heatmap depicting the average
691 expression levels per cluster of the top differentially expressed markers in each cluster. **(G)**
692 Violin plots of selected markers identified by differential expression analysis comparing the
693 two goblet subsets to each other. **(H)** Violin plots of selected markers identified by differential
694 expression analysis of ciliated 1 versus ciliated 2 clusters. **(I)** Dot plot depicting gene
695 expression levels and percentage of cells expressing genes associated with specific lung
696 phenotypes according to the Online Mendelian Inheritance in Man (OMIM) database. Only
697 genes present in the top 50 (per cluster) of our list of differentially expressed genes are
698 depicted in (I). All the differential expression analysis were performed using Wilcoxon rank
699 sum test in Seurat⁴⁸.
700

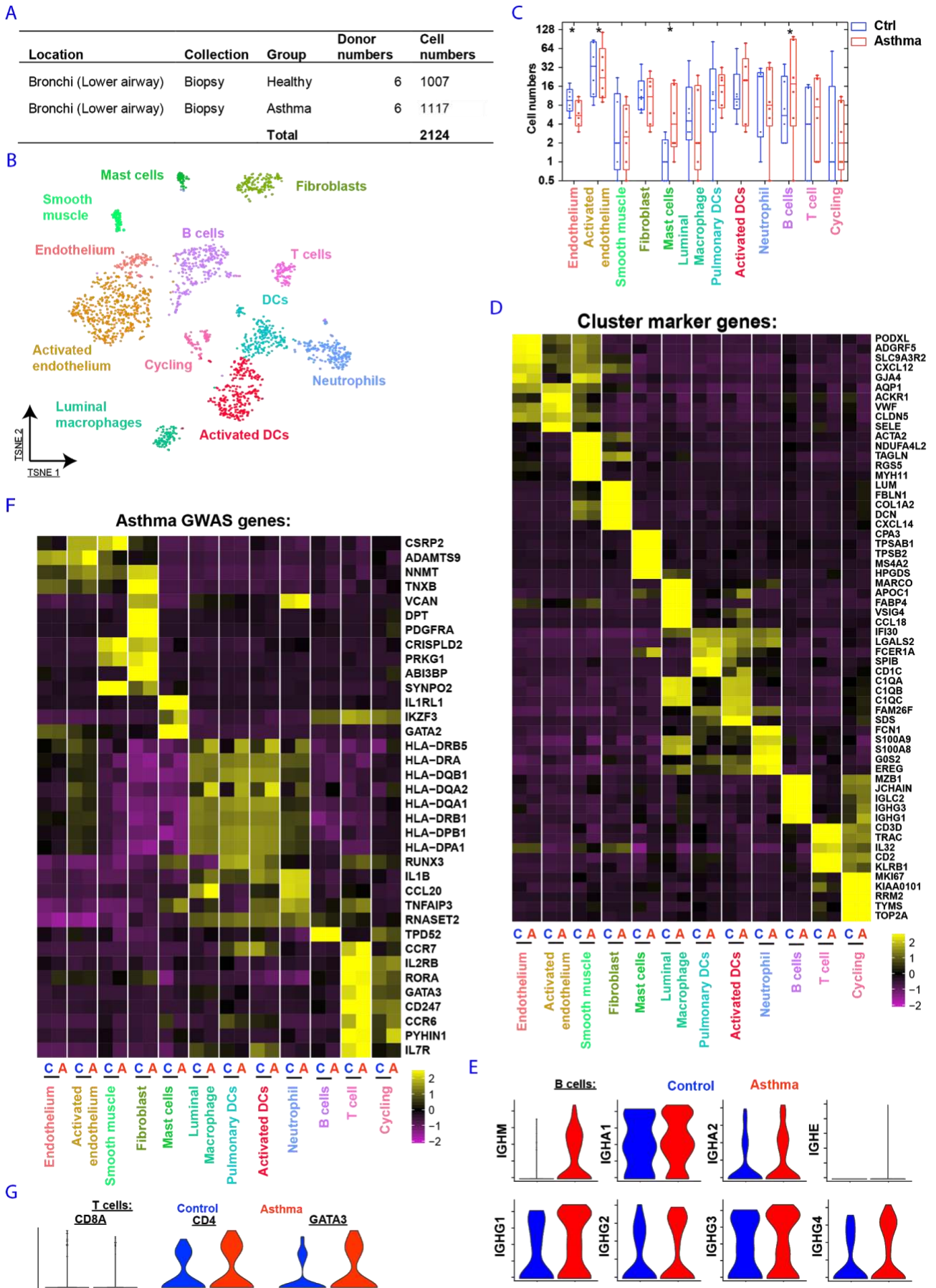


701
 702 **Figure 2. A cellular and molecular map of the stromal and immune components of**
 703 **across the upper and lower human respiratory airways. (A)** Table with details of
 704 anatomical region, tissue source, donors and cell numbers present in this figure. **(B)**
 705 tSNE displaying the major immune and mesenchymal clusters present in the full extent of the
 706 human respiratory tree. **(C)** Pie charts depicting the cellular composition of immune cells by
 707 anatomical region. **(D)** Pie charts depicting the cellular composition of stromal cells in lower
 708 airway biopsies and parenchyma tissue. **(E)** Heatmap depicting the average expression
 709 levels per cluster of the top differentially expressed markers in each cluster. **(F)** Dot plot
 710 depicting gene expression levels and percentage of cells expressing genes associated with
 711 lung phenotypes according to the Online Mendelian Inheritance in Man (OMIM) database.
 712 Only genes present in the top 50 (per cluster) of our list of differentially expressed genes are
 713 depicted in (F). All the differential expression analysis was performed using Wilcoxon rank
 714 sum test in Seurat.



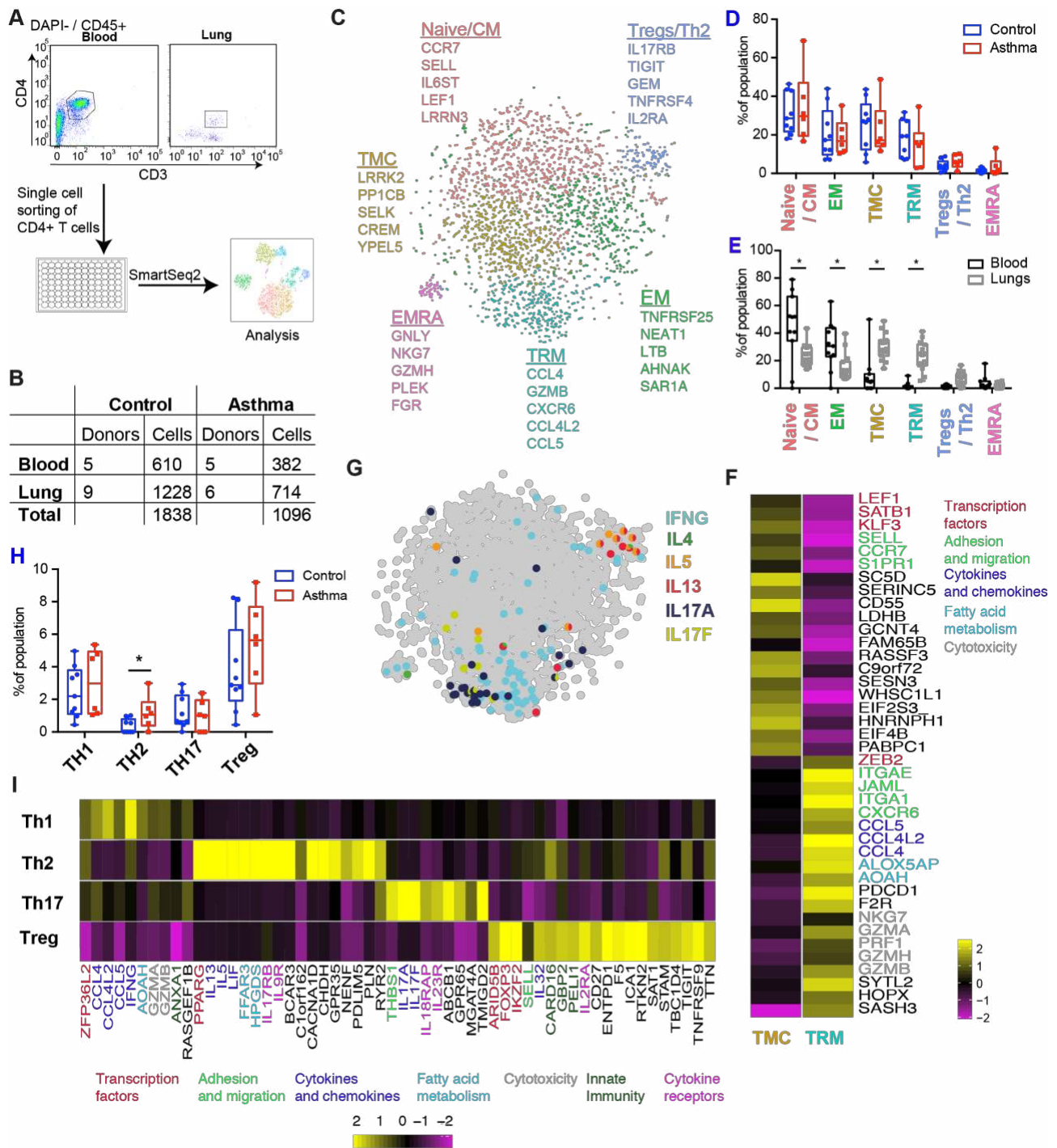
716 **Figure 3. Distinct programs of epithelial cell differentiation in asthmatic versus**
717 **healthy airways. (A)** Table with overview and cell numbers for control and asthma
718 volunteers analysed in this figure. **(B)** tSNE displaying all epithelial cells analysed coloured
719 by their specific cluster assignment. **(C)** Box and whisker plots depicting cell numbers of
720 control and asthma patients in each cluster. **(D)** Heatmap displaying the top five differentially
721 expressed genes per cluster. **(E)** Pseudotime developmental trajectory analysis from
722 Monocle2 depicting how each of the basal, secretory and ciliated subsets relate to each
723 other. **(F)** Binned pseudotime analysis displaying how each subset is ordered in a one-
724 dimensional continuous space. **(G)** Heatmap displaying the expression of asthma genes
725 from GWAS. Only genes present in our list of differentially expressed genes are depicted
726 for each cell cluster. Significance analysed using Fisher's exact test corrected for multiple
727 comparison using the Bonferroni method. Significance calculated using all the clusters
728 present in figures 3 and 4, which were derived from the same set of samples. *represents
729 p-value<0.001. n=6 controls and n=6 asthma. The differential expression analysis used for
730 input in D and E was performed using Wilcoxon rank sum test in Seurat.
731

732



733

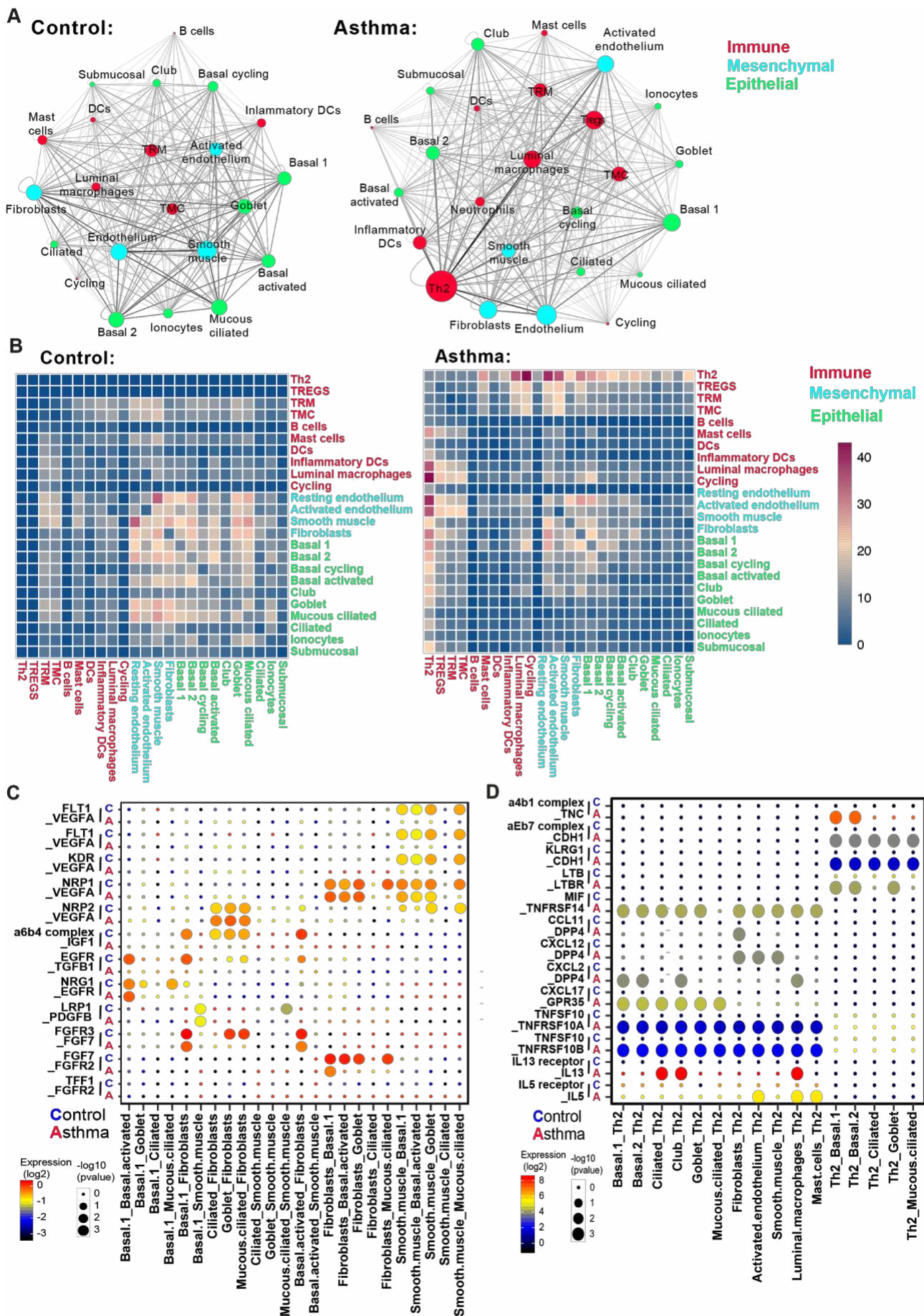
734 **Figure 4. Remodelling of the stromal and Immune compartments in asthmatic**
735 **airways. (A)** Table with the number of donors and cells per volunteer group included in this
736 figure. **(B)** tSNE depicting the immune and stromal cell types identified in the human airway
737 combined dataset of healthy and asthmatic patients. **(C)** Box and whisker plots depicting the
738 cell numbers of healthy and asthmatic cells in each cluster. **(D)** Heatmap displaying gene
739 expression levels of the top 5 differentially expressed genes per cluster. **(E)** Violin plots
740 depicting expression of immunoglobulin genes in B cells. **(F)** Heatmap displaying asthma
741 GWAS gene expression per cluster. Only genes present in the top 50 (per cluster) of our list
742 of differentially expressed genes are shown. **(G)** Violin plots of selected T cell markers in
743 asthma patients. Significance calculated using all the clusters present in figures 3 and 4,
744 which were derived from the same set of samples. *represents p-value<0.001. n=6 controls
745 and n=6 asthma.
746
747
748



749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762

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 control and asthma groups. (C) tSNE displaying clusters of T cells identified by analysing the combined cells from blood and lung from control and asthma groups. (D) Box and whisker plots showing the cluster cell distributions from control and asthma patients. (E) Box and whisker plots depicting the cluster composition per donor according to the tissue source from which the cells were isolated. (F) Heatmap showing the average expression per cluster of genes differentially expressed between the two lung specific CD4 T cell populations. Gene names coloured according to functional categories. (G) tSNE 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 Tregs identified by unbiased clustering. (I) Heatmap of average cluster gene expression of markers differentially

763 expressed between Th1, Th2, Th17 and Treg cells. Gene names coloured according to
764 functional categories. Error bars in **(D)**, **(E)** and **(H)** represent standard deviation.
765 Significance analysed using Multiple t-tests corrected for multiple comparison using the
766 Holm-Sidak method. * represents p-value<0.05. Patient numbers per tissue depicted in (B).
767
768
769



770
771
772

Figure 6. Asthma is characterized by unique cell-to-cell signalling networks. We quantified the predicted cell interactions in healthy and asthmatic airways between all the

773 epithelial and non-epithelial cell clusters identified in figures 3 and 4, plus the lung airway
774 enriched populations of CD4 T cells (Th2, Treg, TMC and TRM) **(A)** Networks depicting cell
775 types as nodes and interactions as edges. Size of cell type proportional to the total number
776 of interactions of each cell type and edge thickness proportional to the number of interactions
777 between the connecting types. **(B)** Heatmap depicting the number of all possible interactions
778 between the clusters analysed. Cell types grouped by broad lineage (epithelia,
779 mesenchymal or immune). **(C)** Dot plot depicting selected epithelial-epithelial and epithelial-
780 mesenchymal interactions enriched in healthy airways but absent in asthma. **(D)** Dot plot
781 depicting selected epithelial-immune and mesenchymal-immune interactions highly
782 enriched in asthma but absent in healthy airways.

783 **Methods**

784

785 **Patient recruitment and ethical approval**

786 Bronchoscopy biopsy (10x and smartseq2 analysis): cohort inclusion criteria for all subjects
 787 were: age between 40 – 65 years and history of smoking <10 pack years. For the asthmatics,
 788 inclusion criteria were: age of onset of asthmatic symptoms ≤12 years, documented history
 789 of asthma, use of inhaled corticosteroids with(out) β2-agonists due to respiratory symptoms
 790 and a positive provocation test (i.e. PC₂₀ methacholine ≤8mg/ml with 2-minute protocol). For
 791 the non-asthmatic controls, the following criteria were essential for inclusion: absent history
 792 of asthma, no use of asthma-related medication, a negative provocation test (i.e. PC₂₀
 793 methacholine >8 mg/ml and adenosine 5'-monophosphate >320 mg/ml with 2-minute
 794 protocol), no pulmonary obstruction (i.e. FEV₁/FVC ≥70%) and absence of lung function
 795 impairment (i.e. FEV₁ ≥80% predicted).

796 Asthmatics stopped inhaled corticosteroid use 6 weeks prior to all tests. All subjects were
 797 clinically characterised with pulmonary function and provocation tests, blood samples were
 798 drawn, and finally subjects underwent a bronchoscopy under sedation. If a subject
 799 developed upper respiratory symptoms, bronchoscopy was postponed for ≥6 weeks.

800 Fiberoptic bronchoscopy was performed using a standardised protocol during conscious
 801 sedation [1]. Six macroscopically adequate endobronchial biopsies were collected for this
 802 study, located between the 3rd and 6th generation of the right lower and middle lobe.
 803 Extracted biopsies were processed directly thereafter, with a maximum of one hour delay.
 804 The medical ethics committee of the Groningen University Medical Center Groningen
 805 approved the study, and all subjects gave their written informed consent. Detailed patient
 806 information below:
 807

Donor	Gender	Age	Classification	Packyears	BMI	FEV1% pred	FEV1/FVC ratio	PC20 methacholine threshold (mg/ml)	10x	Smartseq2
ARMS004	F	57	Ctrl	0	21	99	82	Not reached	no	yes
ARMS005	F	44	Ctrl	4	24	111	82	Not reached	no	yes
ARMS009	M	49	Asthma	0	22	87	69	0.211	no	yes
ARMS014	F	65	Ctrl	0	18	109	78	Not reached	no	yes
ARMS015	M	49	Asthma	0	24	105	67	1.155	no	yes
ARMS018	M	54	Ctrl	3	25	107	80	Not reached	no	yes
ARMS019	F	56	Asthma	0	45	44	55	FEV1 too low	no	yes
ARMS022	M	48	Ctrl	1	30	131	84	Not reached	no	yes
ARMS024	F	51	Asthma	0	33	96	74	0.88	yes	yes
ARMS026	M	64	Ctrl	5	32	103	81	Not reached	yes	yes
ARMS032	M	56	Ctrl	0	22	120	75	Not reached	yes	yes
ARMS033	M	51	Asthma	0	23	98	69	2.63	yes	yes
ARMS035	M	61	Asthma	0	20	89	59	6.71	yes	yes
ARMS038	M	59	Ctrl	0	23	148	74	Not reached	yes	yes
ARMS040	M	51	Ctrl	0	23	119	73	Not reached	yes	yes
ARMS043	F	50	Ctrl	1	23	110	71	Not reached	yes	no
ARMS048	M	53	PersA	0	29	77	54	0.205	yes	no
ARMS050	M	64	PersA	0	21	85	66	0.692	yes	no
ARMS051	M	64	PersA	0	25	59	54	0.631	yes	no
ARMS054	M	60	Ctrl	0	29	116	81	Not reached	yes	no

808

809

810

811 Lung resection (Dropseq analysis): Fresh resected human lung tissue (parenchymal lung
 812 and distal airway specimens) was obtained via the CPC BioArchive at the Comprehensive
 813 Pneumology Center Munich (CPC-M, Munich, Germany). In total, we analysed parenchymal
 814 tissue of uninvolved areas of tumour resection material from four patients. All participants
 815 gave written informed consent and the study was approved by the local ethics committee of
 816 the Ludwig-Maximilians University of Munich, Germany.

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

821 Donor information:

822

ID	Sample	Gender	Age	Smoker	Segment	COPD
muc3843	ASK428	m	62	yes, >100py	lower lobe, left	COPD Gold II
muc4658	ASK440	m	58	yes, 45py	upper lobe, left	no
muc5103	ASK452	m	58	no	main bronchus	no
muc5104	ASK454	m	69	yes, >60py	upper lobe, left	no

823 Lung transplant tissue (10x analysis): Human lung tissue was obtained from deceased organ
824 donors from whom organs were being retrieved for transplantation. Informed consent for the
825 use of tissue was obtained from the donors' families (REC reference: 15/EE/0152 NRES
826 Committee East of England - Cambridge South).

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

831

832 Donor information:

833

834 *Donor 284C*

835 Gender: Male

836 Age band: 55-60

837 BMI: 25.83

838 Cause of Death: hypoxic brain damage

839 Smoking history: smoked 20/day for 25 years

840 Stopped: 08/2000

841 Respiratory related information: Chest X-ray normal on admission. No pleural effusion or
842 pneumothorax. Not diagnosed with asthma, but inhalers for possible seasonal wheeze.
843 Family report only using inhaler approximately 5 times a year. No recent peak flow on record
844 last one in 2008 when it was 460, predicted is 611.

845 Time from death to cell lysis: 12h

846

847 *Donor 290B*

848 Gender: Female

849 Age band: 60-65

850 BMI: 27.77

851 Cause of Death: hypoxic brain damage

852 Smoking history: smoked 15/day for 7 years

853 Stopped: no details

854 Respiratory related information: Respiratory tests all normal on admission; maintaining own
855 airway. GP notes report Acute bronchitis in 1994.

856 Time from death to cell lysis: 2h 27min

857

858 *Donor: 292B*

859 Gender: Male

860 Age band: 55-60

861 BMI: 27.44

862 Cause of Death: Intracranial haemorrhage

863 Smoking history: smoked 20/day for 46 years
864 Stopped: no details
865 Respiratory related information: Chest X-ray normal on admission, lungs appear clear.
866 Bronchoscopy results show global inflamed mucosa. No other history of respiratory issues.
867 Time from death to cell lysis: 18h 50min
868
869 *Donor: 296C*
870 Gender: Female
871 Age band: 30-35
872 BMI: 20.9
873 Cause of Death: Intracranial haemorrhage
874 Smoking history: smoked 20/day for 19 years
875 Stopped: no details
876 Respiratory related information: Chest X-ray shows collapsed left lobe on admission due to
877 consolidation. Right lobe looks normal. No history or record of respiratory issues.
878 Time from death to cell lysis: 15h 30min
879
880 *Donor: 298C*
881 Gender: Male
882 Age band: 50-55
883 BMI: 24
884 Cause of Death: Intracranial haemorrhage
885 Smoking history: not available
886 Stopped: no details
887 Respiratory related information: no details
888 Time from death to cell lysis: 15h 30min
889
890 *Donor: 302C*
891 Gender: Male
892 Age band: 40-45
893 BMI: 34.33
894 Cause of Death: Known or suspected suicide
895 Smoking history: smoked 20/day for 25 years
896 Stopped: no details
897 Respiratory related information: Chest X-ray shows reduced volume in right lung due to
898 collapsed right lower lobe on admission. No history or record of respiratory issues.
899 Time from death to cell lysis: 13h 30min
900
901 Archived formalin-fixed paraffin-embedded (FFPE) lung blocks: Left-over frozen peripheral
902 lung tissues from 6 current smokers and 4 non-smokers who underwent lung resection
903 surgery. These subjects did not have a history of lung disease, apart from lung cancer for
904 which the patients underwent surgery. Lung tissue samples were taken as distant from the
905 tumor as possible. Thus, any possible effect of the tumor on the lung tissue was minimized.
906 All samples were obtained according to national and local ethical guidelines and the
907 research code of the University Medical Center Groningen.
908

Bio sample code	Age	Gender	Smoking status	Diagnosis
GRNG-LNG_MRK085	45	female	current smoker	LCC
GRNG-LNG_MRK125	54	female	current smoker	AC

GRNG-LNG_MRK177	52	female	current smoker	AC
GRNG-LNG_MRK188	70	male	current smoker	SCC
GRNG-LNG_MRK218	61	male	current smoker	Metastasis
GRNG-LNG_MRK245	65	female	current smoker	LCC
GRNG-LNG_0758	81	female	non-smoker	AC
GRNG-LNG_0764	49	male	non-smoker	carcinoid
GRNG-LNG_0802	50	male	non-smoker	metastasis
GRNG-LNG_0830	45	male	non-smoker	IMT

LCC = large cell carcinoma, AC = adenocarcinoma, SCC = squamous cell carcinoma and IMT = inflammatory myofibroblast tumor

909

910

Blood processing

911

912

913

914

915

916

917

Lung tissue processing

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

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 1mg/ml collagenase D and 0.1mg/ml DNase I (Roche) in HBSS (Lonza). This was then placed at 37°C for 1hr with gentle agitation. The single cell suspension was forced through a 70µm 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 scRNAseq 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, 1-1.5 g of tissue was homogenized by mincing with scissors into smaller pieces (~0.5 mm²/piece). Prior to 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 40µm strainer. The bloody filtrate was discarded. The tissue was transferred into 8 ml of enzyme mix consisting of dispase (50 caseinolytic U/ml), collagenase (2 mg/ml), elastase (1 mg/ml), and DNase (30 µg/ml) for mild enzymatic digestion for 1 hour 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 70µm strainer and centrifuged at 300g for 5 minutes at 4°C. The cell pellet was resuspended in 3 ml of red blood cell lysis buffer and incubated at room temperature for 2 minutes 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 minutes 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. 250,000 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/µl.

945 *Rejected lung transplant:* for each sample, 1-2g of tissue was divided in 5 smaller pieces
946 then transferred to 5ml eppendorfs containing 1.5ml 0.5mg/ml collagenase D and 0.1mg/ml
947 DNase I (Sigma) in RPMI. Samples were then finely minced using scissors. Minced tissue
948 was then transferred to a petry dish and extra digestion medium added to completely cover
949 the tissue. Samples were incubated 30min at 37°C. Cells were then passed up and down
950 through a 16-gauge needle 10 times. Samples were incubated for an additional 15min at
951 37°C. Cells were filtered a 70um filter, then spun down for 6min 1400RPM. 1 ml of red blood
952 cell lysis (eBioscience) was added to the pellet during 5min. Cells were resuspended in
953 RPMI + 10%FCS and counted. Dead cells were removed using the Dead Cell Removal Kit
954 (Miltenyi Biotec). In brief, cells were incubated with anti-Annexin V beads for 15min. The cell
955 suspension was then passed through a magnetic column and dead Annexin V+ cells
956 remained in the column, while live cells were collected. Viability was then estimated *via*
957 trypan blue. More than 99% of cells were viable.

958

959 **Flow cytometry**

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

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

966

967 **Cell Sorting**

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

972

973 **Immunohistochemical staining:**

974 Human lung tissue containing large airways were collected from archival formalin-fixed
975 paraffin-embedded (FFPE) blocks (n=10, 6 smokers and 4 non-smokers). Serial sections
976 (~4 µm) were cut for immunohistochemistry (IHC) and immunofluorescent (IF) staining.

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

981 IHC and IF staining was performed as described previously^{49,50}. For the IHC staining cells
982 were stained with a primary antibody (see below for Ab details) and visualized with
983 diaminobenzidine (DAB, Sigma) solution. For the IF staining, cells were stained with primary
984 antibody. Secondary antibodies conjugated to fluorophores (donkey anti rabbit-488, donkey
985 anti mouse-555) were used at a dilution of 1:100. DAPI, dissolved in Dako Fluorescence
986 Mounting Medium (Dako S3023) at a dilution of 1:1000, was used as a nuclear stain.

987

988

989 **Antibody list:**

990

Antibody	Fluorochrome	Clone	Supplier	Catalog
CD45	BB515	HI30	BD biosciences	564585
CD3	PerCP Cy5.5	SP34-2	BD biosciences	552852

CD4	APC Cy7	RPA-T4	BD biosciences	557871
FOXI1	unlabelled	2B8	LSBio	LS-C336930
CFTR	unlabelled	polyclonal	Human atlas	HPA021939
Synaptophysin	unlabelled	SP11	Ventana Medical Systems	790-4407
MUC5AC	unlabelled	45m1	Abcam	ab3649
KRT5	unlabelled	EP1601Y/LL002	Ventana Medical Systems	760-4939
α -Tubulin	unlabelled	DM1A	Sigma/MERCK	T9026-100UL
Donkey anti rabbit-488	AF 488	Polyclonal	Thermo Scientific Fisher	# A-21206
Donkey anti mouse-555	AF 555	Polyclonal	Thermo Scientific Fisher	# A-31570
DAPI	nuclear stain	n/a	Thermo Scientific Fisher	# D3571

991

992

Chromium 10x Genomics library and sequencing

993

Airway biopsy: Single cell suspensions were manually counted using a haemocytometer and concentration adjusted to a minimum of 300 cells/ul. Cells were loaded according to standard protocol of the Chromium single cell 3' kit in order to capture between 2000-5000 cells/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 in order to obtain coverage of at least mean coverage of 100.000 reads/cell.

994

995

996

997

998

999

Lung transplant: Single cell suspensions were manually counted using a haemocytometer and concentration adjusted to 1000 cells/ul. Cells were loaded according to standard protocol of the Chromium single cell 3' kit in order to capture between 2000-5000 cells/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 in order to obtain coverage of at least mean coverage of 100.000 reads/cell.

1000

1001

1002

1003

1004

1005

1006

1007

SmartSeq 2 library preparation and sequencing

1008

1009

1010

1011

1012

1013

1014

Library preparation was performed with minor modifications from the published SmartSeq2 protocol⁵¹. In short, single cells were flow sorted onto individual wells of 96 or 384 wells containing 4ul (96 wells) or 1ul (384 wells) of lysis buffer (0.3% triton plus DNTPs and OligoDT). After sorting, plates were frozen and stored at -80 until further processing. RT, PCR (25 cycles) and nextera library preparation performed as described in ⁵¹.

1015

Dropseq library preparation and sequencing

1016

1017

1018

1019

Drop-seq experiments were performed largely as described previously⁷ with few adaptations during the single cell library preparation. Briefly, using a microfluidic polydimethylsiloxane (PDMS) device (Nanoshift), single cells (100/ μ l) from the lung cell suspension were co-encapsulated in droplets with barcoded beads (120/ μ l, purchased from ChemGenes

1020 Corporation, Wilmington, MA) at rates of 4000 μ l/hr. Droplet emulsions were collected for 15
1021 min/each prior to droplet breakage by perfluorooctanol (Sigma-Aldrich). After breakage,
1022 beads were harvested and the hybridized mRNA transcripts reverse transcribed (Maxima
1023 RT, Thermo Fisher). Unused primers were removed by the addition of exonuclease I (New
1024 England Biolabs), following which beads were washed, counted, and aliquoted for pre-
1025 amplification (2000 beads/reaction, equals \sim 100 cells/reaction) with 12 PCR cycles (primers,
1026 chemistry, and cycle conditions identical to those previously described. PCR products were
1027 pooled and purified twice by 0.6x clean-up beads (CleanNA). Prior to tagmentation, cDNA
1028 samples were loaded on a DNA High Sensitivity Chip on the 2100 Bioanalyzer (Agilent) to
1029 ensure transcript integrity, purity, and amount. For each sample, 1 ng of pre-amplified cDNA
1030 from an estimated 1000 cells was tagmented by Nextera XT (Illumina) with a custom P5
1031 primer (Integrated DNA Technologies). Single cell libraries were sequenced in a 100 bp
1032 paired-end run on the Illumina HiSeq4000 using 0.2 nM denatured sample and 5% PhiX
1033 spike-in. For priming of read 1, 0.5 μ M Read1CustSeqB (primer sequence:
1034 GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC) was used.

1035

1036 **Bulk Transcriptome**

1037 Biopsies were fresh frozen in liquid nitrogen and stored in -80. RNA was extracted after a
1038 few weeks using a combination of Trizol & the RNeasy MinElute Clean Up kit from Qiagen.
1039 RNA was prepared from sequencing using the TruSeq RNA Library Prep Kit v2. Samples
1040 were then sequenced in a HiSeq 4000.

1041

1042

1043 **Single-cell RNA sequencing data alignment**

1044 For SmartSeq2 raw sequencing data, paired-end reads were mapped to the Human genome
1045 (GRCh38) using GSNAP with default parameters⁵². Then, uniquely mapped reads were
1046 counted using htseq-count (<http://www-huber.embl.de/users/anders/HTSeq/>). Low-quality
1047 cells were filtered out using the outlier detection algorithm in R Scater package based on a
1048 cut-off of 2^*MAD (median-absolute-deviation).

1049 10X Genomics raw sequencing data was processed using CellRanger software version
1050 2.0.2 and the 10X human genome GRCh38 1.2.0 release as the reference.

1051 The Dropseq core computational pipeline was used for processing next generation
1052 sequencing reads of the Dropseq scRNA-seq data, as previously described⁷. Briefly, STAR
1053 (version 2.5.2a) was used for mapping⁵³. Reads were aligned to the human reference
1054 genome hg19 (provided by Dropseq group, GSE63269).

1055

1056 **Bulk transcriptome computational analysis**

1057 The bulk samples were aligned using STAR 2.5.1b, using the STAR index from the GRCh38
1058 reference that was used when mapping 10X data, and quantified using HTSeq. The data
1059 was then processed using the Seurat-inspired workflow within Scanpy, adding a number of
1060 "pseudo-bulks" obtained by taking 10X data from donors matching the bulk samples and
1061 summing expression across all cells.

1062

1063 **Data QC**

1064 *General strategy for 10x datasets:* Optimal tissue dissociation conditions are cell-type
1065 dependent, resulting in a certain degree of cell lysis when working with a mixed tissue
1066 sample. This results in substantial background levels of ambient RNA in the single-cell
1067 suspension that vary with cell type composition, so we applied SoupX for background
1068 correction (see below). We analysed each donor sample separately and excluded cells with
1069 a number of genes higher than the median+2SDs for that donor. We further excluded cell
1070 with high number of UMIs and high percentage of mitochondrial reads (see below).

1071 In parallel, we used scrublet (see below) to infer the number of the doublets in the dataset
1072 before applying the filters previously described and excluded any remaining cells predicted
1073 to be doublets that were still present in the dataset. We normalised and scaled our data (see
1074 below), performed clustering (see below) and identified and subset the data into epithelial
1075 and non-epithelial cell groups (as shown in supplementary figures 1 and 6). After separation
1076 between epithelial and non-epithelial, we clustered the cells and performed curated doublet
1077 removal (see below) based on known lineage restricted markers.

1078
1079 *General strategy for Dropseq data:* We normalised and scaled the data, then performed
1080 filtering based on the number of genes and percentage of mitochondrial reads.

1081
1082 *General strategy smartseq2 data:* We normalised and scaled the data, then performed
1083 filtering based on the number of genes and percentage of mitochondrial reads. In order to
1084 avoid potential batch effects from the lung digestion protocol, we corrected the gene
1085 expression of the CD4 SmartSeq2 dataset using a small subset of genes the expression of
1086 which has been recently shown to be highly responsive to enzymatic digestion⁵⁴ : FOS,
1087 ZFP36, JUN, FOSB, HSPA1A, JUNB, EGR1, UBC.

1088 1089 **Ambient RNA correction (SoupX)**

1090 Different batches can be affected by different levels of ambient RNA. To take this into
1091 account, we used the recently developed SoupX method⁵⁵. Briefly, ambient RNA expression
1092 is estimated from the empty droplet pool (10 UMI or less). Expression of these genes in each
1093 cell is then calculated and compared to their proportion in the ambient RNA profile.
1094 Transcripts with a bimodal profile (i.e. that characterize specific groups of cells but are also
1095 highly abundant in empty droplets) are then grouped based on their function. The
1096 contamination fraction derived from the expression of these genes is then used to calculate
1097 the fraction of each droplet's expression corresponding to the actual cell. Finally, this fraction
1098 and the ambient profiles are subtracted from the real expression values.

1099 1100 **UMI and number of genes filtering**

1101 *10x data (After SoupX correction):*
1102 nUMI: minimum 1000/ maximum 60000.
1103 percent.mito, minimum 0 / maximum= 3%

1104 *SmartSeq2 data:*
1105 nGene : minimum 1000 / maximum 4000.
1106 percent.mito, minimum 0 / maximum= 15%

1107 *Dropseq data:*
1108 nGene: minimum 200/ maximum 4000.
1109 percent.mito, minimum 0 / maximum= 20%

1110 1111 **Scrublet**

1112 We used Scrublet (Wolock et al, BioRxiv, <https://doi.org/10.1101/357368>) for unbiased
1113 computational doublet inference. Doublets were identified in each 10X sample individually
1114 using scrublet, setting the expected doublet rate to 0.03 and keeping all other parameters
1115 at their default values. Cells were excluded when they had a score higher than 0.1 for upper
1116 and lower airway samples or higher than 0.05 for parenchyma samples.

1117 1118 **Normalisation and scaling**

1119 Downstream analyses including, normalisation, scaling, clustering of cells and identifying
1120 cluster marker genes were performed using the R software package Seurat⁴⁸ version 2.1
1121 (<https://github.com/satijalab/seurat>).

1122 Samples were log normalised and scaled for the number of genes, number of UMIs and
1123 percentage of mitochondrial reads. The epithelial biopsy dataset comparing healthy and
1124 asthma was also scaled for XIST expression, as we observed some gender specific clusters
1125 of cells that shared lineage markers with the other observed clusters.

1126

1127 **Curated doublet removal**

1128 We combined literature knowledge about cell lineages with over clustering to identify
1129 clusters enriched in potential doublets. The strategy for each dataset is shown below:

1130

1131 *Lung atlas epithelial dataset* (Figure 1 and associated extended data figures): We removed
1132 cells with expression level higher than 0.5 for any of the following markers: PTPRC
1133 (immune), FCER1G (immune), PDGFRA (fibroblast) or PECAM1 (endothelial).

1134

1135 *Lung atlas non-epithelial dataset* (Figure 2 and associated extended data figures): We
1136 removed cells with expression level higher than 0.5 for any of the following markers: EPCAM
1137 (epithelial), KRT5 (basal), "FOXJ1"(ciliated) or MUC5AC (secretory). We then performed first
1138 clustering round (7 PCs, resolution 2) and excluded clusters that expressed combinations
1139 of the following lineage specific markers: MARCO(macrophage), CCL21 (lymphatic
1140 endothelial), TPSB2 (mast cell) or CD3D(T cell). We performed a second clustering round
1141 and exclude a cluster formed by cells from one donor that had low expression TPSB2, while
1142 lacking markers for all other immune lineages.

1143

1144 *Asthma biopsy epithelial cells* (Figure 3 and associated extended data figures): due to the
1145 smaller number of cells, we only performed cluster-based doublet exclusion, without cell
1146 filtering. We performed one round of clustering and removed one clusters with high
1147 expression of PECAM1 (endothelial marker).

1148

1149 *Asthma biopsy non-epithelial cells* (Figure 4 and associated extended data figures): we
1150 performed three rounds of clustering where we excluded clusters with high levels of EPCAM
1151 or KRT5 expressed in much higher levels than immune lineage markers.

1152

1153 **Dimensionality reduction**

1154 We performed PCA dimensionality reduction with the highly variable genes as input. We
1155 then used the PCs to calculate t-Distributed Stochastic Neighbour Embedding (**t-SNE**) for
1156 each dataset, using a perplexity value of 50.

1157

1158 **Data clustering**

1159 We used the function "FindClusters" from Seurat. In brief, this method uses a shared nearest
1160 neighbour (SNN) modularity optimization-based clustering algorithm to identify clusters of
1161 cells based on their PCs. Before constructing the SNN graph, this function calculates k-
1162 nearest neighbours (we used k=30) and then it constructs the SNN graph. The number of
1163 PCs used for each clustering round was dataset dependent and they were estimated by the
1164 elbow of a PCA scree plot, in combination to manual exploration of the top genes from each
1165 PC.

1166

1167 **DE analysis**

1168 We used a Wilcoxon rank sum test to identify differentially expressed genes in all the
1169 comparisons here discussed.

1170

1171 **MatchScore**

1172 We used MatchScore⁸ to quantify the overlap of cell type marker signatures between
1173 experiments, which is based on the Jaccard index. Only marker genes with adjusted p-value
1174 < 0.1 and average log fold change > 1 were considered.

1175

1176 **CellPhoneDB**

1177 We developed a manually curated repository of ligands, receptors and their interactions
1178 called CellPhoneDB (www.cellphonedb.org; Vento-Tormo, Efremova et al., *Nature*, 2018),
1179 integrated with a statistical framework for predicting cell-cell communication networks from
1180 single cell transcriptome data. Briefly, the method infers potential receptor-ligand
1181 interactions based on expression of a receptor by one cell type and a ligand by another cell
1182 type. Only receptors and ligands expressed in more than 30% of the cells in the specific
1183 cluster were considered. In order to identify the most relevant interactions between cell
1184 types, the method prioritizes ligand-receptor interactions that have cell type-specific
1185 expression. To this end, pairwise cluster-cluster interaction analysis are performed by
1186 randomly permuting the cluster labels of each cell 1000 times. For each permutation, the
1187 total mean of the average receptor expression level of a cluster and the average ligand
1188 expression level of the interacting cluster is calculated, and a null distribution is derived for
1189 each receptor-ligand pair in each cluster-cluster interaction. An empirical *p*-value is
1190 calculated from the proportion of the means which are "as or more extreme" than the actual
1191 mean. For the multi-subunit heteromeric complexes, the member of the complex with the
1192 minimum average expression is used for calculating the mean.

1193 Network visualization was done using Cytoscape (version 3.5.1). All the interaction pairs
1194 with collagens were removed from the analysis. The networks layout was set to force-
1195 directed layout.

1196

1197 **Trajectory analysis**

1198 Trajectory analysis was performed using Monocle version 2.2.0²³. We ordered the cells onto
1199 a pseudotime trajectory based on the union of highly variable genes obtained from all cells,
1200 as well as those from only healthy or asthmatic donors.

1201

1202 **Supervised analyses using GWAS genes**

1203 Asthma-associated GWAS gene list was collected using the GWAS Catalog of EMBL-EBI
1204 searching for the term asthma (<https://www.ebi.ac.uk/gwas/>). The list was downloaded on
1205 8th of February 2018. We took the genes that are in the top 50 hits of our single-cell DE
1206 marker list (either epithelial or non-epithelial) and asthma-associated GWAS list (the
1207 "matched" gene list). We then hierarchically clustered the expression matrix of the matched
1208 gene list along its rows (genes) and columns (single cells) and represented this as a
1209 heatmap.

1210

1211 **Neuroendocrine cell identification**

1212 Neuroendocrine cells were identified by the expression of CHGA. Any cell expressing any
1213 amount of CHGA was classified as a neuroendocrine cell.

1214

1215 **OMIM search for lung diseases**

1216 We searched the clinical synopses with known molecular basis in the Online Mendelian
1217 Inheritance in Man (OMIM) database[®] for the following terms: 'pulm*' or 'bronchi*' or 'alveol*'
1218 or 'surfactant' and retrieved 337 entries. These terms were chosen to minimise the return of
1219 genetic conditions causing respiratory insufficiency as a consequence of neuromuscular
1220 dysfunction, skeletal dysplasia (small rib cage) or lung segmentation defects arising in early
1221 embryogenesis. These 337 entries were then manually curated to identify those conditions
1222 with features affecting the bronchial tree, alveoli, lung parenchyma and pulmonary
1223 vasculature. On manual review, entries containing terms such as 'alveolar ridge' of the jaw

1224 and 'pulmonary valve stenosis' and 'pulmonary embolism', but no terms related to primary
1225 pulmonary disorders, were excluded from further consideration. Syndromes caused by
1226 chromosomal disorder or contiguous gene deletion were excluded.

1227

1228 **Statistical methods**

1229 For 10x samples comparing healthy versus asthma, we used Fisher's exact test corrected
1230 for multiple testing with Bonferroni method. Normalised CD4 cluster proportions were
1231 analysed via paired t-tests corrected for multiple testing with Holm-Sidak method.

1232

1233

