Title: Multi-tissue transcriptomics delineates the diversity of airway T cell functions in asthma

Authors: Akul Singhania[1,#], Joshua C Wallington[1], Caroline G Smith[1,2], Daniel Horowitz[3], Karl J Staples[1], Peter H Howarth[1,2], Stephan D Gadola[1,4], Ratko Djukanović[1,2], Christopher H Woelk[1,†,*], Timothy SC Hinks[1,2,5*]

Affiliations:

1. Clinical and Experimental Sciences, University of Southampton Faculty of Medicine, Southampton University Hospital, Southampton, SO16 6YD, UK
2. NIHR Southampton Respiratory Biomedical Research Unit, Southampton University Hospital, Southampton, SO16 6YD, UK
3. Janssen Research and Development, High Wycombe, Buckinghamshire, UK
4. Roche, F. Hoffman-La Roche AG, Konzern-Hauptsitz, Grenzacherstrasse 124, CH-4070, Basel, Switzerland
5. Respiratory Medicine Unit, NDM Experimental Medicine, University of Oxford, Level 7, John Radcliffe Hospital, Oxford, OX3 9DU

* These authors contributed equally to this work

[#]Current address for Akul Singhania: Laboratory of Immunoregulation and Infection, The Francis Crick Institute, London, NW1 1AT, UK.

[†]Current address for Christopher H Woelk: Merck Exploratory Science Center, Merck Research Laboratories, Cambridge MA 02141, USA.

Corresponding author: Dr Timothy SC Hinks, Respiratory Medicine Unit, NDM Experimental Medicine, University of Oxford, Level 7, John Radcliffe Hospital, Oxford, OX3 9DU timothy.hinks@ndm.ox.ac.uk; +44 1865 220885.

Running title: Multi-tissue transcriptomics in asthma

Key words: asthma, transcriptomics, epithelial brushings, sputum, bronchoalveolar lavage

List of abbreviations

BAL: Bronchoalveolar lavage; Benjamini-Hochberg (BH); FACS: Fluorescence activated cell sorting; FDR: False discovery rate; GO: Gene Ontology; ICS: Inhaled corticosteroids; KEGG: Kyoto Encyclopedia of Genes and Genomes; MAIT: Mucosal associated invariant T; RT-qPCR: Real-time quantitative PCR

Grants and Funding

AS was supported by the Faculty of Medicine, University of Southampton, UK. TSCH was supported by Wellcome Trust Research Fellowships (088365/z/09/z and 104553/z/14/z) and by the Academy of Medical Sciences. KJS was supported by a project grant from Asthma UK (08/026).

Authors' contributions

AS performed the microarray data analysis, carried out the RT-qPCR experiment and analysis and was involved in the study design and drafting the manuscript. JCW performed the experiments to characterize the surface

expression of CD14 on subsets of CD3+ cells and appraised the manuscript. CGS provided nursing support for recruitment and sample collection. DH performed microarray hybridization. KJS supervised the CD3+ cell subset experiment and edited the manuscript. PHH, SDG and RD contributed to project conception, design and data analysis. CHW was involved in the data analysis, study design, project supervision and editing the manuscript. TSCH performed patient recruitment, sample collection and processing and was involved in the data analysis, study design and conceptualization, project supervision and drafting the manuscript. All authors approved the final manuscript.

ABSTRACT

Asthma arises from the complex interplay of inflammatory pathways in diverse cell types and tissues. We sought to undertake a comprehensive transcriptomic assessment of the epithelium and airway T cells that remain understudied in asthma, and investigate interactions between multiple cells and tissues. Epithelial brushings and flow-sorted CD3+ T cells from sputum and bronchoalveolar lavage were obtained from healthy subjects (N=19) and asthmatic patients (mild, moderate and severe asthma; N=46). Gene expression was assessed using Affymetrix HT HG-U133+ PM GeneChips and results were validated by real-time quantitative PCR. In the epithelium, IL-13 response genes (*POSTN, SERPINB2*, *CLCA1*), mast cell mediators (*CPA3*, *TPSAB1*), inducible nitric oxide synthase and cystatins (*CST1, CST2*, *CST4*) were upregulated in mild asthma but, except for cystatins, were suppressed by corticosteroids in moderate asthma. In severe asthma – with predominantly neutrophilic phenotype – several distinct processes were upregulated including neutrophilia (*TCN1, MMP9*), mucins and oxidative stress responses. The majority of the disease signature was evident in sputum T cells in severe asthma, where 267 genes were differentially regulated compared to health, highlighting compartmentalisation of inflammation. This signature included IL-17-inducible chemokines (*CXCL1*, *CXCL2*, *CXCL3*, *IL8*, *CSF3*) and chemoattractants for neutrophils (*IL8, CCL3, LGALS3*), T cells and monocytes. A protein interaction network in severe asthma highlighted signatures of responses to bacterial infections across tissues (*CEACAM5*, *CD14*, *TLR2*) including toll-like receptor signalling. In conclusion, the activation of innate immune pathways in the airways suggests that activated T

cells may be driving neutrophilic inflammation and steroid-insensitive IL-17 response in severe asthma.

INTRODUCTION

Asthma is a complex, chronic inflammatory disease of the airways, characterized by inflammation involving the interplay of a multitude of cell types including mast cells, eosinophils, B cells, neutrophils, and airway smooth muscle cells, in interaction with the epithelium (1, 2). Airway T cells are believed to play a central role in driving these diverse pathobiological processes (3). The critical role of Th2 cells, producing type 2 cytokines has long been appreciated, but there is an emerging recognition of a wide variety of novel T cell subsets and innate-like lymphocytes producing a diversity of cytokines, many of which may be steroid-resistant (3, 4). Indeed, whilst the majority of asthma patients respond to regular low-dose inhaled corticosteroids, 5-10% of asthmatics suffer from severe disease, with persistent symptoms despite anti-inflammatory therapy, and often characterized by neutrophilic inflammation (5, 6).

Multi-parameter flow-cytometry has enabled detailed characterization of T cell subsets in the airways according to expression of surface makers and cytokine secretion (3). However, the range of markers that can be studied simultaneously is limited, and results will be restricted by the *a priori* selection of markers and hypotheses to be tested. Therefore a powerful complementary approach is the application of unbiased, hypothesis- generating ‘omics technologies. Several studies in asthma have utilized transcriptomics analysis to detect asthmatic gene signatures in different tissues such as epithelium (7-9), whole sputum (10) and bronchial biopsies (11). However, both an analysis

of the transcriptome of airway T cells, and a multi-tissue investigation of the interaction between these different cells are lacking.

Our objective was to undertake a gene expression analysis of epithelial brushings and airway T cells from sputum and bronchoalveolar lavage, to study the interactions between these tissues, across a range of human asthma. We therefore capitalized on both the unique capacity of flow-cytometry to sort cells with high precision, and the power of an unbiased transcriptomic approach to synthesise a global view of the multiple, diverse pathobiological processes orchestrated by T cells in asthma. Results were validated in a subset of the patients by real-time quantitative PCR.

MATERIALS AND METHODS

Study participants

65 participants (18-70 years) were enrolled from the Wessex Severe Asthma Cohort, NIHR Southampton Respiratory Biomedical Research Unit and outpatient clinics at University Hospital Southampton as previously described (3): 19 healthy non-atopic participants, 15 mild asthma patients on β2-agonists alone, 17 moderate asthmatics on inhaled corticosteroids (ICS) and 14 severe asthmatics with persistent symptoms despite high-dose ICS (n=14) and oral corticosteroids (n=4) (**Table 1**) classified on enrolment applying criteria used previously (3) (see online supplement). The study was approved by the Southampton and South West Hampshire Research Ethics Committee B. All participants provided informed consent.

136

137 **Sample collection**

138 Lung samples were obtained by hypertonic-saline sputum induction,
139 bronchoscopy, bronchoalveolar lavage (BAL) and endobronchial epithelial
140 brushing as previously described (3) (see online supplement). Live CD3+ T
141 cells were isolated from BAL and sputum samples by flow-cytometry using a
142 nine-colour FACSAria^{TM} cell sorter (BD Biosciences)(see online **Figure E1**).

143

144 **RNA isolation and microarray hybridization**

145 RNA was extracted using the Absolutely RNA Nanoprep Kit and quality
146 assessed using Bioanalyzer 2100 (Agilent Ltd., UK). RNA was reverse
147 transcribed, amplified by *in vitro* transcription with the Ovation Pico WTA
148 system V2 (NuGEN Technologies, San Carlos, USA) and hybridized to
149 Affymetrix HT HG-U133+ PM GeneChips (Affymetrix, California, USA) by
150 Janssen Research & Development (Pennsylvania). Gene expression data are
151 available at Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo)
152 under accession number GSE89809.

153

154 **Microarray analysis**

155 Raw microarray gene expression data was normalized using RMA (12) and
156 subjected to quality control procedures as previously described (9). Low
157 expressed genes and genes with low standard deviation across samples were
158 removed from further analysis (further details in online supplement).
159 Differentially expressed genes were identified using *limma* (13) in R (version
160 3.2.2) and subjected to functional enrichment analysis. A protein-protein

interaction network was created using the STRING database (version 9.1) (14) and visualized in Cytoscape (version 2.8.3). Deconvolution analysis to extract cell type-specific profiles was performed using CTen (15) and quantification of relative levels of distinct cell types on a per sample basis for sputum severe asthma samples was carried out using CIBERSORT and the LM22 gene signature database (16).

RT-qPCR analysis

RT-qPCR was performed in a subset of asthma patients and healthy subjects to confirm microarray gene expression, as described in the online supplement.

Statistical analyses

For microarray analysis, a linear mixed modelling approach was used for differential gene expression taking into account pairing of samples, and age and gender differences between groups. Only genes with false discovery rate (FDR) $p<0.05$ corrected for multiple testing using the Benjamini-Hochberg (BH) method (17) were considered significant. For GO analysis, FDR corrected hypergeometric $p<0.05$ were collapsed using REVIGO, where $p<0.05$ (represented as $-\text{Log}_{10}$ p-value) was considered significant. For cell-type deconvolution analysis using CTen, an enrichment score > 2 was considered significant. The cell signatures obtained using the LM22 database in CIBERSORT were compared using a two-tailed student's t-test and an FDR corrected *p-value*<0.05 was considered significant. The IL-13 and IL-17

inducible chemokine signatures were compared by one-way ANOVA in Prism 6.0, with $p<0.05$ considered significant.

RESULTS

Evaluation of gene expression signature across compartments and the disease spectrum

The overall gene expression profile of all samples was evaluated using principal component analysis. CD3+ T cells displayed a distinct profile from epithelial brushings, but not between BAL and sputum sources, primarily reflecting cell type differences (see online **Figure E2**). Moreover, the smoking status of the subjects did not have a significant effect on the overall gene expression profile, with former smokers and current smokers clustering with the non-smokers (see online **Figure E3**). Smokers were thus retained for further analysis. Differentially expressed genes were identified between mild, moderate and severe asthmatics, in turn, and healthy controls within each tissue site (**Figure 1, Table 2**, see online **Tables E2** through **E7**). The majority of the differential gene expression signal was evident in severe asthma patients, primarily in CD3+ T cells isolated from sputum (**Table 2**), highlighting the inadequacy of current corticosteroid therapy in reversing disease related gene expression changes in sputum.

Mild and moderate asthma

In the epithelium in mild asthma, IL-13 response genes (*POSTN*, *SERPINB2* and *CLCA1*) and proteases that are products of mast cells (*CPA3*, *TPSAB1*),

as well as inducible nitric oxide synthase (*NOS2*) were upregulated compared with health (**Figure 1A,** online **Table E2**). These responses, which are corticosteroid responsive (7), were not significant in moderate asthmatics suggesting that treatment is effective in clearing this disease signature in moderate ICS-treated asthmatics. Moreover, the stress/stimulus-response genes (*SCNN1G, LTF* and *C4A*) were downregulated in the epithelium in moderate asthmatics (**Figure 1B,** online **Table E3**). In contrast, cystatins (*CST1*, *CST2* and *CST4*) were highly upregulated in both mild and moderate asthmatics compared with health. The differential expression of *SERPINB2*, *POSTN* and *CPA3* in mild asthma and *CST1* in mild and moderate asthma was confirmed by RT-qPCR (**Figure 2A**).

In CD3+ T cells isolated from sputum, cellular processes such as GO-annotated '*cell cycle processes*' were upregulated in both mild and moderate asthmatics compared with healthy controls (**Figure 1D** and **1E,** online **Tables E5** and **E6**). As activation induces quiescent T cells to leave cell cycle arrest, this suggests there is increased activation of airway T cells in asthma. However, '*immune response*' was observed in moderate asthmatics only, showing the significant effect of disease in this group of patients in the sputum compartment. In CD3+ T cells isolated from BAL there were no genes in mild asthmatics and only one gene (*SCGB1A1*) in moderate asthmatics that was differentially expressed compared to healthy controls suggesting a lesser impact of disease on gene expression in this compartment, and is consistent with sputum being more representative of inflammation in the proximal airways whilst BAL samples more distal compartments, which we have

recently shown to have different transcriptomic profiles (9). These results indicate the significant impact of asthma in inducing disease related gene expression changes in the epithelium and CD3+ T cells isolated from sputum.

Severe asthma

The large numbers of differentially expressed genes identified when comparing severe asthma to healthy controls (**Table 2, Figure 1**) were combined across compartments (epithelium, sputum and BAL, n=293) to understand the interaction of genes between these cell types and their contribution to disease. GO analysis was performed with terms being assigned arbitrary colours (**Figure 3A**) ('*Defense response*' and '*immune system process*' were assigned the same colour due to significant overlap in genes). Gene signatures typical of responses to bacterial infection were prominent as evidenced by GO analysis (**Figure 3A**) and by the upregulation of *CEACAM5* (18) in the epithelium and *CD14* and *TLR2* in sputum T cells (online **Table E4** and **E7**). *Haemophilus influenzae* was identified in the BAL sample from one of the seven severe neutrophilic asthmatic subjects analysed by metagenomics and standard respiratory bacterial culture, detected by both techniques. This patient's symptomatology responded well to targeted antibiotics. Next, we investigated these findings using cell-type enrichment analysis (see online supplementary data, **Figures E4A** and **E4B, Table E8**) and additional flow-cytometric studies to confirm these molecules are indeed expressed on the surface of T cells, being highest on the innate-like mucosal associated invariant T (MAIT) cell subset (online **Figure E4C**).

A protein interaction network was created for the combined list of differentially expressed genes (shown as nodes) demonstrating known direct interactions (shown as edges) between the protein products of these genes in these tissues (**Figure 3B**). Prominent features of the network are recruitment of neutrophils, innate immune activation and responses to bacterial products. Proteins involved in specific GO terms are highlighted in the network with respective colours. The '*toll-like receptor signalling pathway*' (KEGG ID – hsa04620) was also significantly enriched in sputum T cells and the protein products of the genes involved in this pathway are highlighted in the protein interaction network. Differential expression of *CEACAM5*, *TCN1* and *TLR2* was confirmed by RT-qPCR (**Figure 2B**). These results indicate the effect of severe asthma in epithelial brushings and T cells obtained from sputum in activating gene expression linked to innate immune pathways.

IL-13 and IL-17 inducible chemokines

The expression levels of the IL-13 (*POSTN*, *SERPINB2* and *CLCA1*) and IL-17 (*CXCL1*, *CXCL2*, *CXCL3*, *IL8* and *CSF3*) inducible chemokines – established by Choy *et al.* (4) – were plotted for all tissues across all asthma severities relative to their respective healthy controls (**Figure 4**) in order to evaluate Th2 and Th17 signatures. The corticosteroid responsive IL-13 inducible chemokines were highly expressed in epithelial samples obtained from mild asthmatics but diminished with an increase in asthma severity (**Figure 4A**). Expression of these genes did not differ in sputum or BAL T cells. In contrast, the IL-17 inducible chemokines were only modulated in asthma samples obtained from sputum T cells and their relative expression

285 increased with an increase in asthma severity, but not in BAL T cells or
286 epithelium (**Figure 4B**). An additional analysis was carried out in subjects with
287 paired samples from epithelium and sputum (21 asthmatic and 9 healthy
288 controls) with similar results (online **Figure E5**), suggesting that these
289 differences were not a result of non-adherence to therapy. These results
290 suggest that the corticosteroid therapy works effectively against the IL-13
291 inducible chemokines that are primarily modulated in epithelium, whereas
292 better therapies targeted towards the IL-17 inducible chemokines in sputum
293 are needed for the management of severe asthma.
294
295
296 **DISCUSSION**
297 By applying a transcriptomic analysis to investigate inflammatory processes in
298 multiple tissues, across a spectrum of asthma, we have demonstrated a wide
299 variety of inflammatory pathways in which airway T cells are implicated. As
300 visualised in the rich protein interaction network (**Figure 3B**), the striking
301 number of differentially expressed genes in sputum T cells, and the perfect
302 separation of health and asthma by hierarchical clustering of sputum T cell
303 data (**Figures 1D, 1E** and **1F**), our findings highlight the diverse, central roles
304 that T cells play in orchestrating the pathobiology of asthma.
305
306 Using untargeted global gene expression profiling, a robust IL-13 response
307 signature (*CLCA1*, *POSTN*, *SERPINB2*) as well as a mast cell signature was
308 evident in mild asthmatics, providing important validation of previous studies
309 in mild disease (7, 8). There was also a prominent upregulation of cystatins

(*CST1*, *CST2, CST4*) in mild and moderate asthma, with *CST1* being the most upregulated gene in both mild and moderate disease, suggesting an important role for these molecules in asthma. Cystatins are small proteins present in all biological fluids which are potent, reversible, competitive inhibitors of cysteine proteinases (cathepsins) (19) and therefore have recognised roles in inflammation. In addition cystatins have unrelated potent immunomodulatory functions and can, for instance, induce synthesis of TNF-α and IL-10 (20). Recent studies have shown differential expression of *CST1* in airway epithelial cells in exercise-induced bronchoconstriction and identified a single nucleotide polymorphism which controls expression of CST1 and is associated with FEV_1 decline in asthma (21). Others have reported elevation of, cystatin C (*CST3*) in asthma (22) – which we did not observe – and more recently of *CST2* (11). These steroid-sensitive IL-13 and mast cell responses, and nitric oxide synthase were not prevalent in moderate asthmatics, and were abrogated in severe asthma (**Figure 4A**) likely due to high-dose therapeutic corticosteroids. Conversely, all three cystatin genes remained highly upregulated in moderate – though not severe – disease (online **Tables E2**, **E3** and **E4**). Cystatins therefore warrant further investigation as potential novel therapeutic targets.

In contrast to the steroid-sensitive processes dominant in mild and moderate asthma, the IL-17 inducible chemokines (*CXCL1*, *CXCL2*, *CXCL3*, *IL8* and *CSF3*) showed the highest expression in severe asthma (**Figure 4B**), consistent with this pathway being resistant to, or even enhanced (4) by corticosteroids. Interestingly, whilst Choy *et al.* (4) initially described this

specific gene signature in cultured normal human bronchial epithelial cells treated with IL-17A and TNF-α, we did not observe this signature in epithelial cells *in vivo* but rather in sputum T cells. It is interesting that this complex of chemokines is not suppressed by steroids, whilst IL-17A itself is suppressed by dexamethasone (4). As neither IL-17A, nor flow-cytometrically measured airway Th17 cells were elevated in severe asthma in this patient cohort (3), this suggests other cell types and processes – such as TNF-α secretion – may be responsible for the induction of these chemokines *in vivo*.

The protein interaction network (**Figure 3B**) explores the interaction of T cells and epithelium in severe asthma, illustrating the involvement of airway T cells in several key pathobiological processes. For instance, the recruitment of neutrophils in severe asthma may be driven by T cell-derived IL-8, which is prominent within the network, visible from the colour intensity (upregulation) and size (connectivity) of this node. This node interacts closely with other molecules involved with granulocyte recruitment including *CCL3* (C-C Motif Chemokine Ligand 3 / Macrophage Inflammatory Protein 1-α), *LGALS3* (Galectin-3, a regulator of neutrophil and eosinophil recruitment) (23) and *MMP9* (matrix metalloproteinase 9): a type-IV collagenase involved with IL-8 induced mobilization of neutrophils, which we have shown is associated with neutrophilic asthma (6). The particular strength of these signals in our dataset may be related to the specific focus on neutrophilic asthma in this cohort (67% of the severe group).

One important T-cell–epithelial interaction is between T-cell sputum *IL8, LGALS3* and the epithelial upregulation of *MUC5AC* (**Figure 3B**). Woodruff *et al.* (8) have previously described an increase in the ratio of epithelial *MUC5AC*:*MUC5B* expression in mild-moderate Th2 high asthma. We too observed repression of epithelial *MUC5B* in mild asthma, but also saw *MUC5AC* was strongly upregulated in severe asthma, despite high dose corticosteroids and predominantly neutrophilic inflammation, suggesting this mechanism may also be common to steroid-resistant neutrophilic asthma.

A second important set of interactions in the protein interaction network (**Figure 3B**) involves upregulation of epithelial *SNCA* and *ALAS2* with activation of T cell pathways typical of responses to bacterial products including *SLC11A1* (Natural Resistance-Associated Macrophage Protein 1, *NRAMP1*), *RELA, JUN,* and *RIPK2*. Both these epithelial genes are implicated in response to oxidative stress and tobacco smoking. In a recent study of gene-environment interactions, single nucleotide polymorphisms in *SNCA* (α-synuclein) were identified as the most strongly associated with accelerated lung function decline induced by tobacco or air pollution (24). Likewise *ALAS2* expression mediates the association between smoking and production of IL-6 and C-reactive protein (25). In our dataset we did not observe a dramatic effect of smoking on overall gene expression profile. This is likely because only 1 subject was a current smoker so differences in expression profiles between tissues and related to asthma will have dominated the principle component analyses. However, whilst only one participant reported current smoking status, four other severe participants had

384 a significant previous smoking history, thus differential expression of these
385 genes could be related to current unreported smoking, passive exposure, or
386 to persistent activation of processes initiated by previous tobacco exposure.
387 However, these processes could equally be attributed to other sources of
388 oxidative stress in asthma particularly inflammatory cell-derived reactive
389 oxygen species.
390
391 A third prominent feature of the protein interaction network (**Figure 3B**) is the
392 upregulation in sputum T cells of pathways associated with innate immune
393 activation, particularly *TLR2*, *CD14* and *SLC11A1* (*NRAMP1*). The latter is
394 expressed by innate lymphocytes and augments their activation, particularly in
395 the γδT cell subset (26), biasing towards a strong IFN-γ response and Th1
396 rather than Th2 responses (26). Expression of *TLR2* and *CD14* are classically
397 associated with innate immune cells, particularly monocytes (27, 28).
398 However we confirmed *CD14* protein expression by surface staining on T
399 cells, finding the highest expression on the innate-like MAIT cell subset
400 (online **Figure E4C**), a cell type we have recently shown to play an important
401 role in host defense against airways infection by bacteria such as
402 *Haemophilus influenzae* (29). Furthermore increased expression of *TLR2* and
403 *CD14* has been reported on CD8+ T cells in COPD (30, 31), and in whole
404 sputum in non-eosinophilic asthma, associated with increased sputum
405 endotoxin (32). Thus upregulation in severe neutrophilic asthma of such
406 innate pathogen receptors on T cells, together with other components of the
407 NFκB signalling pathway and epithelial upregulation of *CEACAM5* (a receptor
408 exploited by airway pathogens including *Haemophilus influenzae* (18))

409 constitutes a signature suggestive of bacterial infection driving the
410 inflammation of the airways in these individuals.
411
412 Finally it should be noted that our protein interaction network recapitulates
413 many other more well-described features of asthma, including high
414 upregulation of *FCER1G* (the high affinity IgE receptor), *PTGS2*
415 (cyclooxygenase-2), *NOS2* (inducible nitric oxide synthase, the source of
416 increased exhaled nitric oxide in untreated asthma) (1) and *CCL17*
417 (chemotactic to CCR4+ pulmonary T cells) (33) as well as downregulation of
418 *IFNA21* (interferon-α), consistent with the known deficiency in type I interferon
419 production (34). These lends confidence in the validity of the gene expression
420 signals emerging from our dataset and to the potential value in exploring other
421 genes of interest we have identified here. These include *CNTF* (ciliary
422 neurotrophic factor), which has been implicated both in enhancing IgE
423 production in allergic disease (35) and *in vitro* in a response to gastro
424 oesophageal reflux (36); an important comorbidity in severe asthma.
425
426 Our study has some limitations and the results should be interpreted carefully.
427 Firstly, epithelial brushings are of mixed cell-type and investigations in pure
428 isolated cell populations would be ideal. Cell-type enrichment tools could
429 provide additional information, but these tools are still under development and
430 depend on specific marker genes, which are lacking for the cell subsets that
431 comprise epithelial brushings. Moreover, the CIBERSORT analysis must be
432 interpreted with care as it is only an *in silico* analysis to identify relative levels
433 of transcripts in complex gene expression data, and these gene signatures

may in fact be present to different extents in different cell types. Nonetheless, others have demonstrated airway epithelial cells comprise 95-97% of cells obtained by bronchial brushings in patients with airways disease, with proportions of airway macrophages and lymphocytes constant between health and severe disease (37). Secondly our sample size is limited, and it will be interesting to replicate these findings in a future clinical cohort including a greater representation of severe eosinophilic phenotypes. Additionally, there was an imbalance in atopy between cases and controls. A cohort of atopic controls in a future study would help differentiate gene expression changes related to asthma, atopy, or both. Thirdly we would like to have included data on T cells from collagenase-dispersed bronchial biopsies, as perhaps the most clinically relevant tissue site. This was attempted, but cell numbers were limiting providing insufficient RNA for analysis.

In summary, by examining transcriptomic signatures across multiple tissues and cell types we have observed airway T cells orchestrating recruitment of neutrophils, monocytes and T cells, in severe asthma. We have generated important confirmatory evidence of the roles of IL-13- and IL-17-inducible chemokines and direct evidence of dysregulation *in vivo* of genes recently linked to genetic susceptibility to asthma and lung function decline. The innate immune signatures observed in both epithelium and T cells putatively suggest a pathogenic role for bacterial airways infection. Thus our results support future investigation of therapies targeting IL-17-inducible chemokines and bacterial airways colonisation in severe neutrophilic asthma.

Acknowledgements

The authors would like to thank staff of the Southampton Centre for Biomedical Research and the staff of the NIHR Wellcome Trust Southampton Clinical Research Facility and the volunteers who gave of their time and enthusiasm to make this research possible. Current address of AS – Laboratory of Immunoregulation and Infection, The Francis Crick Institute, London, NW1 1AT, UK.

REFERENCES

1. Holgate ST. Pathogenesis of asthma. *Clinical & Experimental Allergy* 2008; 38: 872-897.

2. Brightling CE, Bradding P, Symon FA, Holgate ST, Wardlaw AJ, Pavord ID. Mast-cell infiltration of airway smooth muscle in asthma. *New England Journal of Medicine* 2002; 346: 1699-1705.

3. Hinks TS, Zhou X, Staples KJ, Dimitrov BD, Manta A, Petrossian T, Lum PY, Smith CG, Ward JA, Howarth PH, Walls AF, Gadola SD, Djukanovic R. Innate and adaptive T cells in asthmatic patients: Relationship to severity and disease mechanisms. *Journal of Allergy and Clinical Immunology* 2015; 136: 323-333.

4. Choy DF, Hart KM, Borthwick LA, Shikotra A, Nagarkar DR, Siddiqui S, Jia G, Ohri CM, Doran E, Vannella KM, Butler CA, Hargadon B, Sciurba JC, Gieseck RL, Thompson RW, White S, Abbas AR, Jackman J, Wu LC, Egen JG, Heaney LG, Ramalingam TR, Arron JR, Wynn TA, Bradding P. TH2 and TH17 inflammatory pathways are reciprocally regulated in asthma. *Science Translational Medicine* 2015; 7: 301ra129.

5. Wenzel SE, Szefler SJ, Leung DY, Sloan SI, Rex MD, Martin RJ. Bronchoscopic evaluation of severe asthma: persistent inflammation associated with high dose glucocorticoids. *American Journal of Respiratory and Critical Care Medicine* 1997; 156: 737-743.

6. Hinks TS, Brown T, Lau LC, Rupani H, Barber C, Elliott S, Ward JA, Ono J, Ohta S, Izuhara K, Djukanovic R, Kurukulaaratchy RJ, Chauhan A, Howarth PH. Multidimensional endotyping in patients with severe asthma reveals

492 inflammatory heterogeneity in matrix metalloproteinases and chitinase 3-like
493 protein 1. *Journal of Allergy and Clinical Immunology* 2016; 138: 61-75.
494 7. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly
495 S, Ellwanger A, Sidhu SS, Dao-Pick TP, Pantoja C. Genome-wide profiling
496 identifies epithelial cell genes associated with asthma and with treatment
497 response to corticosteroids. *Proceedings of the National Academy of*
498 *Sciences* 2007; 104: 15858-15863.
499 8. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, Arron
500 JR, Koth LL, Fahy JV. T-helper type 2–driven inflammation defines major
501 subphenotypes of asthma. *American Journal of Respiratory and Critical Care*
502 *Medicine* 2009; 180: 388-395.
503 9. Singhania A, Rupani H, Jayasekera N, Lumb S, Hales P, Gozzard N,
504 Davies DE, Woelk CH, Howarth PH. Altered Epithelial Gene Expression in
505 Peripheral Airways of Severe Asthma. *PloS one* 2017; 12: e0168680.
506 10. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional
507 phenotypes of asthma defined by gene expression profiling of induced sputum
508 samples. *Journal of Allergy and Clinical Immunology* 2011; 127: 153-160.
509 11. Kuo CS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, Hoda U,
510 Rossios C, Sousa A, Wilson SJ, Howarth P, Dahlen B, Dahlen SE, Chanez P,
511 Shaw D, Krug N, Sandstrom T, De Meulder B, Lefaudeux D, Fowler S,
512 Fleming L, Corfield J, Auffray C, Sterk PJ, Djukanovic R, Guo Y, Adcock IM,
513 Chung KF. A Transcriptome-driven Analysis of Epithelial Brushings and
514 Bronchial Biopsies to Define Asthma Phenotypes in U-BIOPRED. *American*
515 *journal of respiratory and critical care medicine* 2016.

516 12. Irizarry RA, Hobbs B, Collin F, Beazer‐Barclay YD, Antonellis KJ, Scherf
517 U, Speed TP. Exploration, normalization, and summaries of high density
518 oligonucleotide array probe level data. *Biostatistics* 2003; 4: 249-264.
519 13. Smyth GK. Limma: linear models for microarray data. Bioinformatics and
520 computational biology solutions using R and Bioconductor: Springer; 2005. p.
521 397-420.
522 14. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-
523 Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP. STRING v10: protein–
524 protein interaction networks, integrated over the tree of life. *Nucleic Acids*
525 *Research* 2014: 1003.
526 15. Shoemaker JE, Lopes TJ, Ghosh S, Matsuoka Y, Kawaoka Y, Kitano H.
527 CTen: a web-based platform for identifying enriched cell types from
528 heterogeneous microarray data. *BMC Genomics* 2012; 13: 460.
529 16. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD,
530 Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue
531 expression profiles. *Nature Methods* 2015; 12: 453-457.
532 17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical
533 and powerful approach to multiple testing. *Journal of the Royal Statistical*
534 *Society Series B (Methodological)* 1995: 289-300.
535 18. Klaile E, Klassert TE, Scheffrahn I, Müller MM, Heinrich A, Heyl KA,
536 Dienemann H, Grünewald C, Bals R, Singer BB. Carcinoembryonic antigen
537 (CEA)-related cell adhesion molecules are co-expressed in the human lung
538 and their expression can be modulated in bronchial epithelial cells by non-
539 typable Haemophilus influenzae, Moraxella catarrhalis, TLR3, and type I and
540 II interferons. *Respiratory research* 2013; 14: 85.

19. Henskens YMC, Veerman ECI, Nieuw Amerongen AV. Cystatins in health and disease. *Biological Chemistry-Hoppe Seyler* 1996; 377: 71-86.

20. Lalmanach G, Saidi A, Marchand-Adam S, Lecaille F, Kasabova M. Cysteine cathepsins and cystatins: from ancillary tasks to prominent status in lung diseases. *Biological Chemistry* 2015; 396: 111-130.

21. Luo W, Obeidat M, Di Narzo AF, Chen R, Sin DD, Pare PD, Hao K. Airway Epithelial Expression Quantitative Trait Loci Reveal Genes Underlying Asthma and Other Airway Diseases. *Am J Respir Cell Mol Biol* 2016; 54: 177-187.

22. Cimerman N, Brguljan PM, Krašovec M, Šuškovič S, Kos J. Serum cystatin C, a potent inhibitor of cysteine proteinases, is elevated in asthmatic patients. *Clinica Chimica Acta* 2000; 300: 83-95.

23. Gao P, Gibson PG, Baines KJ, Yang IA, Upham JW, Reynolds PN, Hodge S, James AL, Jenkins C, Peters MJ, Zhang J, Simpson JL. Anti-inflammatory deficiencies in neutrophilic asthma: reduced galectin-3 and IL-1RA/IL-1beta. *Respiratory Research* 2015; 16: 5.

24. Curjuric I, Imboden M, Nadif R, Kumar A, Schindler C, Haun M, Kronenberg F, Kunzli N, Phuleria H, Postma DS, Russi EW, Rochat T, Demenais F, Probst-Hensch NM. Different genes interact with particulate matter and tobacco smoke exposure in affecting lung function decline in the general population. *PLoS One* 2012; 7: e40175.

25. Huan T, Joehanes R, Schurmann C, Schramm K, Pilling LC, Peters MJ, Magi R, DeMeo D, O'Connor GT, Ferrucci L, Teumer A, Homuth G, Biffar R, Volker U, Herder C, Waldenberger M, Peters A, Zeilinger S, Metspalu A, Hofman A, Uitterlinden AG, Hernandez DG, Singleton AB, Bandinelli S,

566 Munson PJ, Lin H, Benjamin EJ, Esko T, Grabe HJ, Prokisch H, van Meurs
567 JB, Melzer D, Levy D. A Whole-Blood Transcriptome Meta-Analysis Identifies
568 Gene Expression Signatures of Cigarette Smoking. *Human Molecular*
569 *Genetics* 2016.
570 26. Hedges JF, Kimmel E, Snyder DT, Jerome M, Jutila MA. Solute carrier
571 11A1 is expressed by innate lymphocytes and augments their activation.
572 *Journal of Immunology* 2013; 190: 4263-4273.
573 27. Antal-Szalmas P, Strijp JA, Weersink AJ, Verhoef J, Van Kessel KP.
574 Quantitation of surface CD14 on human monocytes and neutrophils. *Journal*
575 *of Leukocyte Biology* 1997; 61: 721-728.
576 28. Sabroe I, Jones EC, Usher LR, Whyte MKB, Dower SK. Toll-like receptor
577 (TLR) 2 and TLR4 in human peripheral blood granulocytes: a critical role for
578 monocytes in leukocyte lipopolysaccharide responses. *The Journal of*
579 *Immunology* 2002; 168: 4701-4710.
580 29. Hinks TS, Wallington JC, Williams AP, Djukanovic R, Staples KJ,
581 Wilkinson TM. Steroid-induced Deficiency of Mucosal-associated Invariant T
582 Cells in the COPD Lung: Implications for NTHi Infection. *American Journal of*
583 *Respiratory and Critical Care Medicine* 2016; 194: 1208-1218.
584 30. Jersmann HPA. Time to abandon dogma: CD14 is expressed by non-
585 myeloid lineage cells. *Immunology and Cell Biology* 2005; 83: 462-467.
586 31. Freeman CM, Martinez FJ, Han MK, Washko GR, McCubbrey AL,
587 Chensue SW, Arenberg DA, Meldrum CA, McCloskey L, Curtis JL. Lung
588 CD8+ T cells in COPD have increased expression of bacterial TLRs.
589 *Respiratory Research* 2013; 14: 1.

590 32. Simpson JL, Grissell TV, Douwes J, Scott RJ, Boyle MJ, Gibson PG.
591 Innate immune activation in neutrophilic asthma and bronchiectasis. *Thorax*
592 2007; 62: 211-218.
593 33. Staples KJ, Hinks TS, Ward JA, Gunn V, Smith C, Djukanovic R.
594 Phenotypic characterization of lung macrophages in asthmatic patients:
595 overexpression of CCL17. *Journal of Allergy and Clinical Immunology* 2012;
596 130: 1404-1412.e1407.
597 34. Wark PA, Johnston SL, Bucchieri F, Powell R, Puddicombe S, Laza-
598 Stanca V, Holgate ST, Davies DE. Asthmatic bronchial epithelial cells have a
599 deficient innate immune response to infection with rhinovirus. *Journal of*
600 *Experimental Medicine* 2005; 201: 937-947.
601 35. Kimata H. Ciliary neurotrophic factor preferentially enhances spontaneous
602 IgE production by B cells from atopic patients. *Neuropeptides* 2004; 38: 92-
603 97.
604 36. Chiu HY, Chen CW, Lin HT, Hsieh CC, Lin SS, Cheng CM. Study of
605 gastric fluid induced cytokine and chemokine expression in airway smooth
606 muscle cells and airway remodeling. *Cytokine* 2011; 56: 726-731.
607 37. Hodge SJ, Hodge GL, Holmes M, Reynolds PN. Flow cytometric
608 characterization of cell populations in bronchoalveolar lavage and bronchial
609 brushings from patients with chronic obstructive pulmonary disease.
610 *Cytometry B Clin Cytom* 2004; 61: 27-34.
611

612 **TABLES**

613 **Table 1. Demographic and clinical characteristics of the**

614 **participants**

615

Variable	Healthy Controls	Mild Asthma	Moderate Asthma	Severe Asthma
n	19	15	17	14
Demographics				
Sex (M/F)	13 / 6	8 / 7	7 / 10	7 / 7
Age (median [range], years)	28 (20-65)	26 (21-64)	33 (21-56)	57 (31-67)
Pulmonary function				
FEV_1 (% predicted)	108 (105-112)	88 (86-102)	100 (90-107)	63 (52-75)
FEV_1 reversibility (%)	4.7 (2.3-7.9)	13 (11-18)	12 (8.9-17)	19 (10-26)
PEFR (% predicted)	102 (94-111)	98 (90-107)	96 (91-99)	69 (58-81)
PEFR variability (%)	0 (0-11)	17 (11-20)	21 (17-29)	14 (11-23)
PD20 (mg methacholine)	Negative	0.19 (0.061-0.59)	0.22 (0.097-0.53)	Not tested
Exhaled nitric oxide (ppb, at 50 L/s)	19 (11-23)	52 (29-106)	27 (14-46)	22 (15-39)
Clinical				
Atopy (Skin test positive, Y/N)	0 / 19	15 / 0	15 / 2	9 / 5
No. of skin test allergens positive	0 (N/A)	6 (4-7)	3 (3-5)	4 (4-6)
Peripheral eosinophil count (10^9/L)	0.1 (0.1-0·2)	0.2 (0.1-0.4)	0.2 (0.2-0.3)	0.2 (0.1-0.2)
Total IgE (IU/ml)	18 (8.0-43)	172 (48-365)	119 (56-188)	84 (19-552)
Body mass index (kg/m^2)	25.8 (23.6-28.8)	23.6 (22.7-26.5)	25 (23.0-32.0)	33.6 (28.3-41.8)
Smoking status				
Never	16	14	13	9
Former (Mean pack-years)	3 (3.7)	1 (5)	4 (1.8)	4 (27)
Current (Mean pack-years)	0	0	0	1 (49)
Duration of asthma (years)	N/A	18 (15-25)	22 (11-24)	43 (15-50)
ACQ score	N/A	0.6 (0.50-1.1)	1.3 (0.71-1.7)	2.8 (2.4-3.4)
GINA level of control (n, %)				
Controlled	N/A	8 (53)	3 (18)	0 (0)
Partly controlled	N/A	6 (40)	11 (65)	0 (0)

Uncontrolled	N/A	1 (6.7)	3 (18)	14 (100)
Treatment				
Inhaled steroids	No	No	Yes	Yes
Dose (equivalent mcg BDP)	N/A	N/A	400 (400-800)	1440 (1280-1920)
Maintenance oral corticosteroids (Y/N)	No	No	No	4 / 10
Mean dose if taken (mg prednisolone/day)		0	0	8.1
Short acting β agonist (Y/N)	No	Yes	Yes	Yes
Long acting β agonist (Y/N)	No	No	7 / 10	14 / 0
Leukotriene receptor antagonist (Y/N)	No	No	0 / 17	11 / 3
Step on GINA treatment algorithm	N/A	1	2 - 3	4 - 5
Inflammatory subtype (n, %)				
Neutrophilic	4 (33)	3 (23)	3 (21)	8 (67)
Eosinophilic	0 (0)	2 (15)	2 (14)	2 (17)
Mixed granulocytic	0 (0)	0 (0)	0 (0)	1 (8.3)
Paucigranulocytic	8 (66)	8 (62)	9 (64)	1 (8.3)
Sputum cell differential (%)				
Macrophages	52 (31-63)	49 (36-62)	53 (31-63)	24 (16-34)
Neutrophils	34 (8.0-64)	34 (24-46)	24 (14-51)	71 (61-79)
Epithelial	4.9 (2.4-21)	4.3 (3.3-9.0)	6.5 (1.5-17)	1.3 (0.13-6.5)
Eosinophils	0.38 (0.0-0.75)	1.5 (0·75-1.8)	1 (0.25-1.3)	0.5 (0.0-2.5)
Lymphocytes	0.25 (0.0-0.81)	0.25 (0.0-0.75)	0 (0.0-0.50)	0 (0.0-0.22)
BAL cell differential (%)				
Macrophages	82 (74-88)	70 (60-77)	81 (75-85)	72 (46-87)
Neutrophils	2.6 (1.9-6.0)	2.5 (1.8-4.5)	3.5 (1.8-5.9)	7.5 (5.0-22)
Epithelial	11 (4.7-17)	21 (15-35)	12 (7.3-18)	9.5 (3.5-11)
Eosinophils	0.5 (0.0-0.69)	2 (1.0-3.0)	1 (0.19-2.6)	0 (0.0-1.0)
Lymphocytes	1.4 (1.0-2.1)	1.5 (0.75-3.0)	1.5 (0.69-2.1)	0.75 (0.0-1.3)
Relevant comorbidities (n, %)				
Allergic rhinitis	0 (0)	12 (80)	8 (47)	5 (36)
Nasal Polyps	0 (0)	0 (0)	1 (5.9)	4 (29)
Eczema	2 (11)	7 (47)	5 (29)	2 (14)
Hypertension	1 (5.2)	0 (0)	2 (12)	3 (21)
Gastro-oesophageal reflux	0 (0)	0 (0)	2 (12)	2 (14)

Other (n=1 each)	Urticaria	Hypothyroidism	Addison's disease, bronchiectasis, multiple sclerosis, benign intracranial hypertension	Neurofibromatosis I, hypothyroidism, psoriasis

Values are medians with interquartile ranges, unless stated otherwise. N/A: not available. The inflammatory subtype is based on sputum differentials using the following cut-points: neutrophilic: >61%, eosinophilic: >3%). Percentages given are derived from those subjects with valid data. Abbreviations are as follows: *ACQ*, asthma control questionnaire; *BAL*, bronchoalveolar lavage; *BDP*, beclometasone dipropionate; *CT*, computed tomograph; FEV_1, pre-bronchodilator forced expiratory volume in 1 second; *FVC*, forced vital capacity; GINA, Global Initiative for Asthma; *mcg*, micrograms; *mg*, milligrams; *PEFR*, peak expiratory flow rate; PD_{20}, provocative dose causing a 20% fall in FEV_1.

635 **Table 2. Numbers of differentially expressed genes**

636

Sample source	Cell type	Differentially expressed genes (compared to healthy controls)		
		Mild Asthma	Moderate Asthma	Severe Asthma
Epithelium	Mixed	14 (↑14; ↓0)	6 (↑3; ↓3)	25 (↑7; ↓18)
BAL	CD3+	0	1 (↑0; ↓1)	1 (↑0; ↓1)
Sputum	CD3+	14 (↑10; ↓4)	22 (↑16; ↓6)	267 (↑166; ↓101)

BAL, bronchoalveolar lavage

In the parenthesis below each number, the number of upregulated genes in indicated, followed by the number of downregulated genes.

637

638

639

640

641

642

643

644

645

646

647

648

649

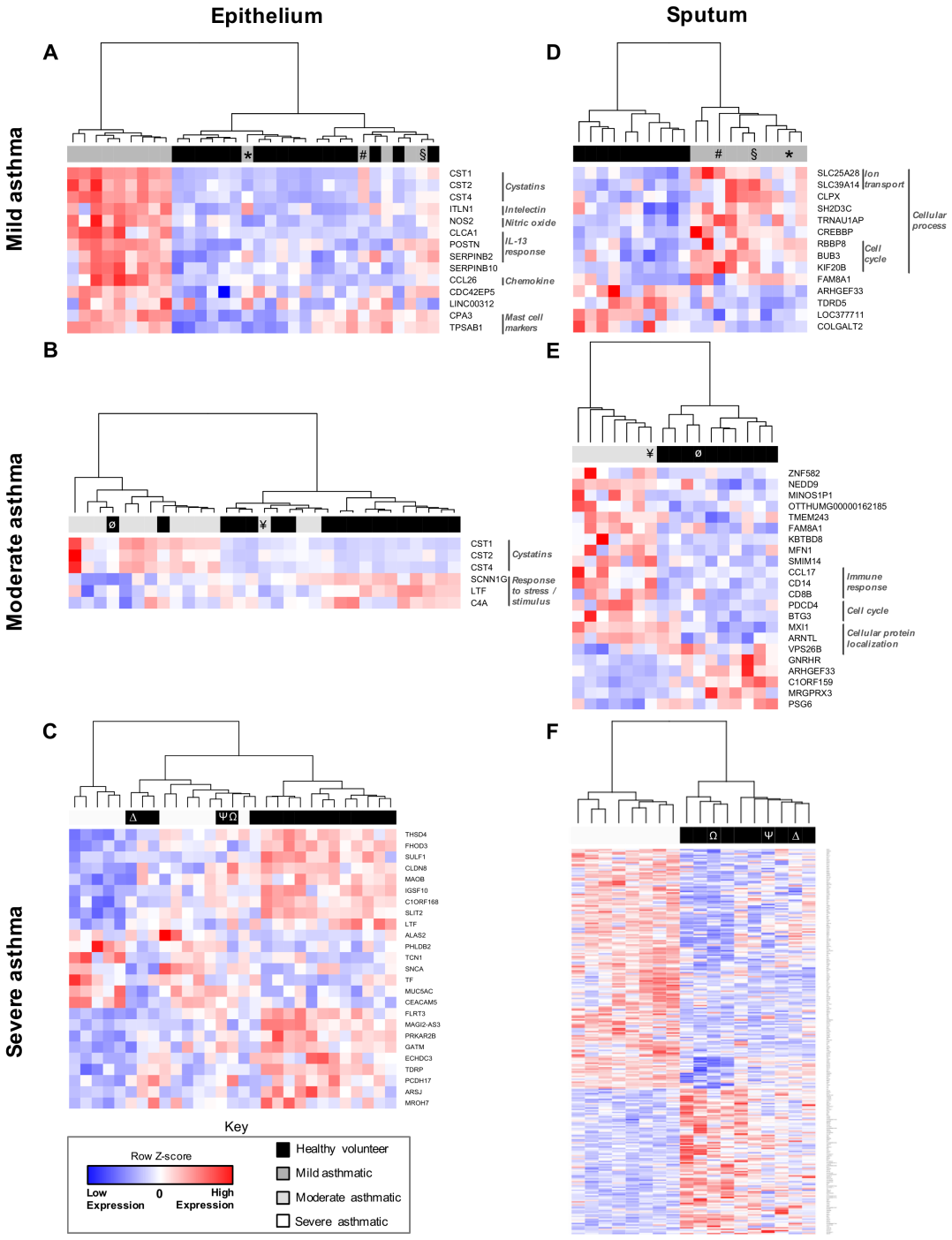
650

651

652

FIGURES

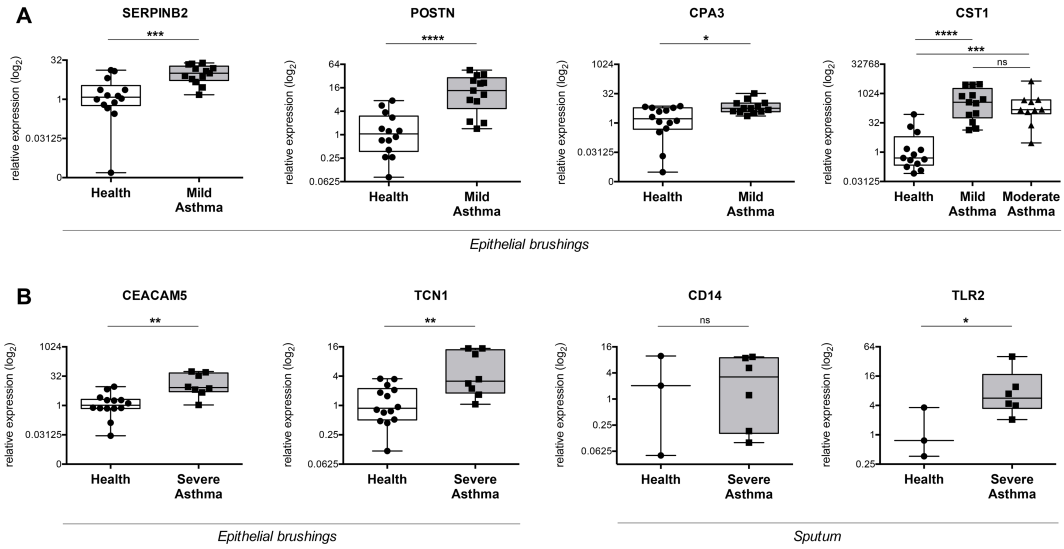
Figure 1.



Heatmaps depicting unsupervised hierarchical clustering of samples and differentially expressed genes in (A) mild, (B) moderate and (C) severe asthma in epithelium and (D) mild, (E) moderate and (F) severe asthma in sputum, compared to healthy controls. Distances were calculated using

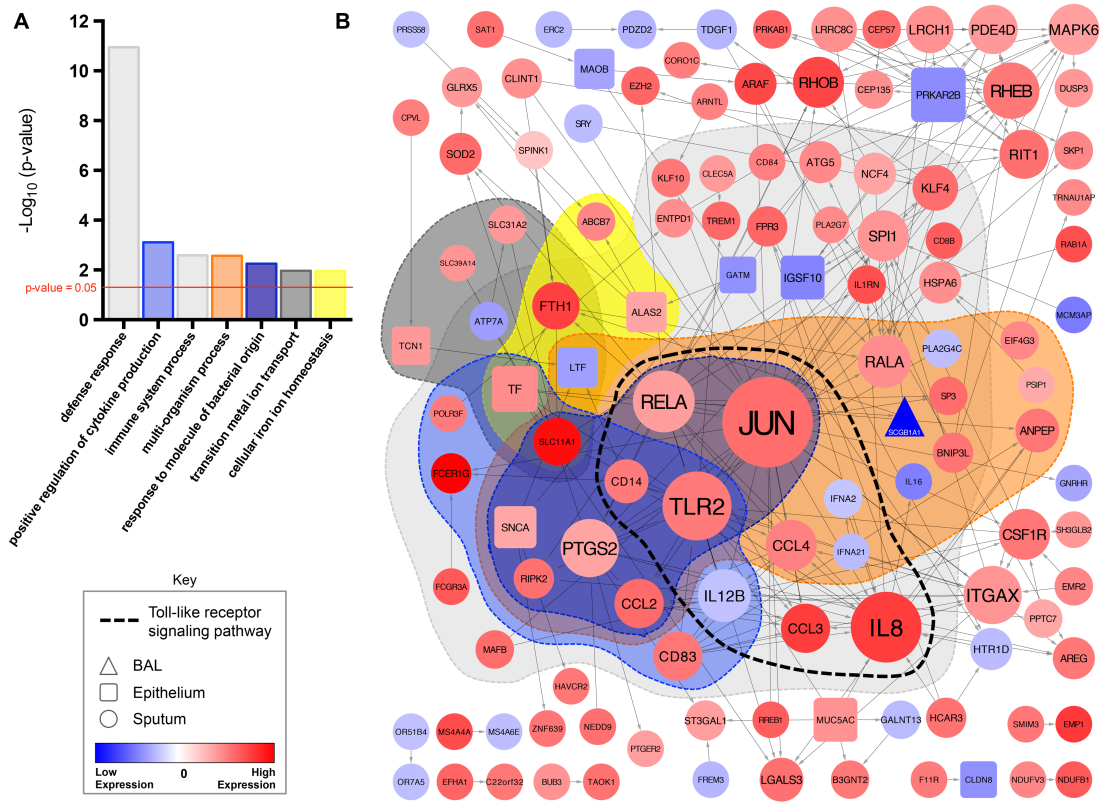
660 Euclidean correlation metric and clustered using Ward's method. Gene
661 expression values were averaged and scaled across the row to indicate the
662 number of standard deviations above (red) or below (blue) the mean, denoted
663 as row Z-score. Colour bars at the top represent healthy volunteers (black)
664 and mild or moderate asthmatics (grey). *, #, §, ø, ¥, ∆, Ψ and Ω represent
665 those samples in epithelium that did not cluster in their respective groups, but
666 had paired samples in sputum that clustered perfectly within the respective
667 groups.
668

Figure 2.



RT-qPCR validation of microarray findings in (A) mild and moderate asthmatics for genes *SERPINB2*, *POSTN*, *CPA3* and *CST1* in epithelial brushings and in (B) severe asthmatics for genes *CEACAM5* and *TCN1* in epithelial brushings, and *CD14* and *TLR2* in sputum. Statistical significance was assessed by Mann-Whitney (two groups) and Kruskal-Wallis with *post-hoc* Dunn's (multiple groups) tests. *, $P \leq .05$; **, $P \leq .01$; ***, $P \leq .001$; ****, $P \leq .0001$, *ns*, not significant.

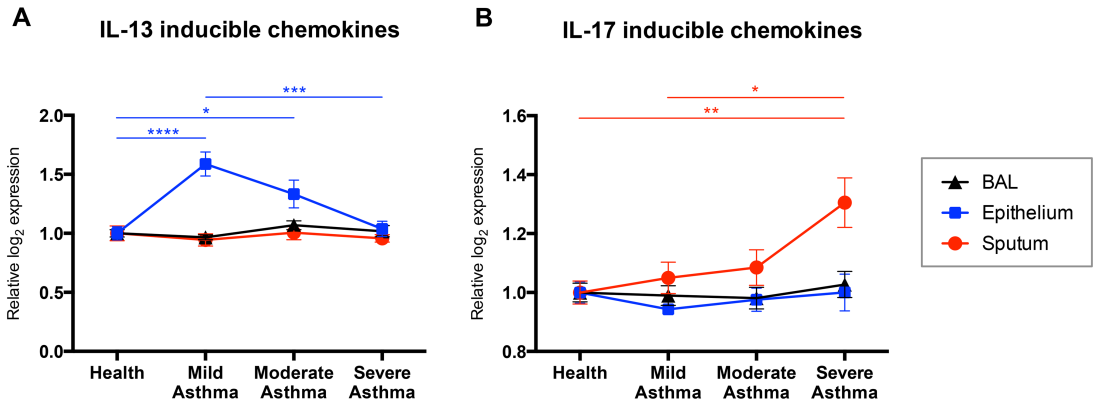
Figure 3.



Functional enrichment of combined differentially expressed gene list in severe asthma across all tissues. (A) Gene ontology analysis with arbitrary colours assigned to biological GO terms. *Defense response* and *immune system process* were assigned the same colour due to a large overlap in genes in these terms. GO terms with FDR corrected $P<0.05$ were collapsed into categories of related terms using REVIGO, where –Log10 p-values were considered significant. A red line indicating $P<0.05$ is shown here. (B) Protein interaction network representing the interactions of the protein products of differentially expressed genes. The size of the node is reflective of the number of interactions and the colour key indicates upregulation (red) or downregulation (blue) compared with healthy controls. Triangles, squares and circles represent BAL, epithelium and sputum, respectively. Functionally related genes are grouped into colours reflecting their respective GO terms

694 from a). Toll-like receptor signaling KEGG pathway (FDR corrected $P<0.05$) is
695 indicated with a dotted line.
696

Figure 4.



Relative gene expression of IL-13 and IL-17-inducible chemokines. The microarray expression levels of the (A) IL-13 (*POSTN*, *SERPINB2* and *CLCA1*) and (B) IL-17 (*CXCL1*, *CXCL2*, *CXCL3*, *IL8* and *CSF3*) inducible chemokines are plotted for all tissues across all asthma severities relative to their respective healthy controls. Black triangles, blue squares and red circles represent BAL, epithelium and sputum, respectively. Statistical significance was assessed by one-way ANOVA, with the colour of asterisks and bars representing the significance in the respective tissues. *, $P \leq .05$; **, $P \leq .01$; ***, $P \leq .001$; ****, $P \leq .0001$.