ASSESSING THE UNIFIED AIRWAY HYPOTHESIS IN CHILDREN VIA TRANSCRIPTIONAL PROFILING OF THE AIRWAY EPITHELIUM

Anthony Kicic (PhD) [1,2,3,4,5*], Emma de Jong (PhD) [1*], Kak-Ming Ling (BSc)[3], Kristy Nichol (BSc)[6,7], Denise Anderson (MBiostat)[1], Peter A.B Wark (MBBS)[6,7], Darryl A. Knight (PhD) [6,7,8], Anthony Bosco (PhD) [1] & Stephen M. Stick (MBBChir) [1,3,4,5] on behalf of WAERP[1] and AusREC[1,9,10]

1: Telethon Kids Institute, Centre for Health Research, The University of Western Australia, Nedlands, 6009, Western Australia, Australia.

2: Occupation and Environment, School of Public Health, Curtin University, Perth, 6845, Western Australia, Australia.

3: School of Biomedical Sciences, The University of Western Australia, Nedlands, 6009, Western Australia, Australia.

4: Department of Respiratory Medicine, Princess Margaret Hospital for Children, Perth, 6001, Western Australia, Australia.

5: Centre for Cell Therapy and Regenerative Medicine, School of Medicine and Pharmacology, The University of Western Australia and Harry Perkins Institute of Medical Research, Nedlands, 6009, Western Australia, Australia.

6: School of Medicine and Public Health, University of Newcastle, Callaghan, New South Wales, Australia.

7: Priority Research Centre for Healthy Lungs, Hunter Medical Research Institute, Newcastle, New South Wales, Australia.

8: Department of Anesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, Canada.

9. Robinson Research Institute, University of Adelaide, Adelaide, SA, Australia

10. Hunter Medical Research Institute, Priority Research Centre for Asthma and Respiratory Disease, New Lambton Heights, NSW, Australia

* authors contributed equally.

Corresponding author: Associate Professor Anthony Kicic, Telethon Kids Institute, Nedlands, Perth, 6009, Western Australia, Australia Ph: (618) 6319 1799. email: Anthony.Kicic@telethonkids.org.au

This article has an online data supplement, which is accessible from this issue's table of content at www.atsjournals.org

FUNDING: This work was supported by a grant from Asthma Australia. SMS is a NHMRC Practitioner Fellow. AK is a Rothwell Family Fellow. AB is supported by a Fellowship from the Simon Lee Foundation.

AUTHOR DECLARATION: All authors declare that there are no conflicts of interest financial or otherwise with this manuscript submission.

AUTHOR CONTRIBUTION: Study conception and design: SMS, AK, PAW, DAK, AB; Sample recruitment and clinical expertise: AK, KML, KN, SMS, PAW; Experimental work: AK, KML, KN, EJ, DA, AB; Analysis and interpretation of data: AK, EJ, SMS, PAW, DAK,

AB, DA; Drafting of manuscript: AK, EJ, SMS, DAK, AB, PAW; Critical revision of manuscript: SMS, AK, EJ, KML, KN, DA, PAW, DAK, AB

TOTAL WORD COUNT: 5051

ABSTRACT (254)

Background: Emerging evidence suggests that disease vulnerability is expressed throughout the airways; the so-called "unified airway hypothesis" but the evidence to support this is predominantly indirect. **Objectives:** To establish the transcriptomic profiles of the upper and lower airway and determine their level of similarity irrespective of airway symptoms (wheeze) and allergy. **Methods:** We performed RNA-sequencing on upper and lower airway epithelial cells from 63 children with or without wheeze and accompanying atopy, utilizing differential gene expression and gene co-expression analyses to determine transcriptional similarity. **Results:** We observed ~91% homology in the expressed genes between the two sites. When co-expressed genes were grouped into modules relating to biological functions, all were found to be conserved between the two regions, resulting in a consensus network containing 16 modules associated with ribosomal function, metabolism, gene expression, mitochondrial activity and anti-viral responses through interferon activity. Although symptom associated gene expression changes were more prominent in the lower airway, they were reflected in nasal epithelium and included; *IL1RL1, PTGS1, CCL26* and *POSTN*. Through network analysis we identified a cluster of co-expressed genes associated with atopic-wheeze in the lower airway, which could equally distinguish atopic and non-atopic phenotypes in upper airway samples. **Conclusions:** We show that the upper and lower airway are significantly conserved in their transcriptional composition, and that variations associated with disease are present in both nasal and tracheal epithelium. Findings from this study supporting a unified airway imply that clinical insight regarding the lower airway in health and disease can be gained from studying the nasal epithelium.

KEY MESSAGES:

- Nasal and tracheal epithelium exhibit similar transcriptional profiles, with some minor differences reflective of site-specific function
- Gene expression changes associated with wheeze and atopy are reflected in both tracheal and nasal epithelium
- These changes include some genes that are well characterised in the context of asthma (*CLCA1*, *IL1RL1*, *CCL26* and *POSTN)*

CAPSULE SUMMARY:

We identified 91% transcriptional similarity between the upper and lower airway supporting the 'unified airway hypothesis'. Furthermore, gene expression changes associated with wheeze and atopy were also conserved in both regions.

KEYWORDS: airway epithelium, unified airway hypothesis, transcriptomics, gene expression, biological processes

ABBREVIATIONS: American Thoracic Society (ATS), Cadherin 10 (CDH10), Cadherin 15 (CDH15), Cadherin 19 (CDH19), Chemokine (C-C motif) Ligand 5 (CCL5), Chemokine (C-C motif) Ligand 26 (CCL26),Chloride Channel Accessory 1 (CLCA1), Chronic Obstructive Pulmonary Disease (COPD), Cystatin 1 (CST1), Cystatin 2 (CST2), Cystatin 4 (CST4), Cystic Fibrosis (CF), Granulocyte-Macrophage Colony Stimulating Factor (GM-CSF),Interleukin (IL), Interleukin 1 Receptor Like 1 (IL1RL1), International Study of Asthma and Allergies in Children (ISAAC), Keratin 1 (KRT1), Keratin 23 (KRT23), Keratin 33A (KRT33A), Keratin 34 (KRT34), Keratin 36 (KRT36), Keratin 37 (KRT37), Keratin 71

(KRT71), Nasal Epithelial Cells, (NEC)Neurotrophic Receptor Tyrosine Kinase 1 (NTRK1), Periostin (POSTN), Prostaglandin-Endoperoxide Synthase 1 (PTGS1), Protocadherin 8 (PCDH8).Protocadherin 11 X-linked (PCDH11X), Radioallergosorbent Test (RAST), Tracheal Epithelial Cells (TEC), Weighted Gene Co-expression Network Analysis (WGCNA),

INTRODUCTION

The 'unified airway hypothesis' postulates that diseases of the upper and lower airway are both manifestations of a single process[1]. Its principle lies in the concept of a continuous respiratory tract lined with respiratory epithelial cells encompassing the nose and middle ear, which extends to the terminal bronchioles[2]. Thus, under disease settings, a pathological process in one region of the airway would affect the function of the entire airway[3]. This is particularly evident in chronic airway diseases such as chronic rhinosinusitis and asthma. These diseases have traditionally been evaluated and treated as separate disorders but advances in the understanding and knowledge of the underlying processes have moved opinion towards the concept of unifying their management[4].

The upper and lower respiratory tracts share many anatomical and histological properties, as well as the airflow into and out of the lungs[5]. They also share a common susceptibility and response to different allergens, infections and pollutants[6-9]. In the setting of asthma, comparable inflammatory cell profiles, particularly in T-cells, mast cells and macrophages have been reported for both the upper and lower respiratory tracts[10, 11]. In addition, similar chemokine and cytokine production have also been reported including interleukin (IL) 5, IL-4, IL-13, Chemokine (C-C motif) ligand 5 (CCL5), and granulocyte-macrophage colony stimulating factor (GM-CSF) levels[12, 13]. In contrast, eosinophil infiltration, epithelial shedding and basement membrane thickness have all been reported to be greater in lower airway of asthmatic patients[10, 14]. Furthermore, the highly-developed vasculature evident in the nose contrasts with the smooth muscle bundles that surround the tracheal/ bronchial airways[15]. This dichotomy highlights the need for additional clarification as to whether there exists a unified airway, especially since the complex interactions between the nose and lung

are not well understood but may have important implications for the pathogenesis of respiratory diseases.

In this study, we implemented a transcriptional profiling-based approach to systematically analyze 63 children with or without atopy and/or wheeze to explore similarities and differences in gene expression between the upper and lower respiratory tracts. Specifically, we tested the hypothesis that the transcriptomic profiles of both the nasal and lower tracheal epithelium would be comparable. Using weighted correlation network analysis on all study participants, we identified 16 modules of co-expressed (and functionally enriched) genes to be conserved between the lower and upper airway. Further, we identified an atopic-wheeze gene signature within tracheal epithelium that was equally able to distinguish between atopic and non-atopic phenotypes within nasal epithelium. These findings provide insight into the intrinsic biological differences inherent to epithelium in the context of airway symptoms and support the concept of a unified airway. Our data suggests that overall, there is conservation of the underlying transcriptional networks between the lower and upper airway epithelium in terms of the biological processes that are active at each site, however some regional specificity in terms of relative gene expression levels is evident.

METHODS

Study population

The study was approved by the Perth Children's Hospital and St John of God Hospital's Human Ethics Committees and written consent was obtained from each participant's legal guardian after being fully informed about the purpose of the study. Airway samples were obtained from 63 children undergoing elective surgery for non-respiratory related conditions (Table 1) and those symptomatic for an existing bacterial or viral chest infection were excluded from the study. Children were identified as having a history of wheeze based on reported chest wheezing on auscultation by each participant's treating physician. All had regular documented episodes of recurrent and persistent wheeze in the preceding year, complemented with a history of documented episodes greater than 12 months prior to recruitment. A smaller number of older children with recurrent wheeze were also diagnosed with asthma which was defined as physician-diagnosed based upon physician documented wheezing episodes in the 12-months preceding their recruitment and confirmed by positive responses on the International Study of Asthma and Allergies in Children (ISAAC) and American Thoracic Society (ATS) respiratory questionnaires[16,17]. Atopy was determined by a positive radioallergosorbent test (RAST) to a designated panel of allergens (house dust mite, mixed grasses, various animal hair, and different moulds). Children were stratified into four groups for sub-analysis including; (1) healthy, (2) atopy no wheeze (atopy alone), (3) wheeze no atopy (wheeze alone) and (4) wheeze and atopy.

Sampling procedure and cell types

Lower Airway Epithelial Cells (AECs) were harvested via trans-laryngeal, non-bronchoscopic brushing of the tracheal mucosa of children via an endotracheal tube as previously

described[18-20]. Matched upper AECs were also obtained and was executed using a rotational movement of the brush against the nasal turbinate. The brush tip was then inserted and cut off into cold sterile media (RPMI- 1640) containing 20% heat inactivated fetal calf serum (FCS). The process was repeated at least once more before the samples were transported back to the laboratory on ice for immediate processing. An unsheathed soft nylon cytology brush (BC 25105; Olympus) was used to sample both regions and performed by the same physician for all participants to minimize sample variance and potential study confounders.

RNA isolation and sequencing

Collection tubes, containing nasal or tracheal AECs were vortexed for a period to disrupt the cells off the brush and the larger cell clumps. The brushes are removed and placed into another collection tube which is then vortexed for a period to release any remaining cells. The collection media was then pooled, centrifuged at 500 x g for 7 minutes at 4°C, and resuspended in 5 mL of collection media. A cell count was then performed and approximately one million AECs were fractionated out, pelleted and resuspended in 350 µL RLT buffer containing 1% (v/v) ß-mercaptoethanol (QIAGEN) and stored at -80°C until all samples were obtained. RNA was extracted using the Ambion Purelink® RNA mini kit (Thermo Scientific) per the manufacturer's instructions and appropriate RNA samples shipped to the Australian Genome Research Facility for library preparation (TruSeq Stranded mRNA Library Prep Kit, Illumina) and sequencing (Illumina HiSeq2500, 50-bp single-end reads, v4 chemistry). RNA integrity was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies); median RNA integrity number was 8.3 (range 7.8–9). The RNA-seq data have been deposited in NCBI's Gene Expression Omnibus[21], and are accessible through GEO Series accession number GSE118761 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118761).

Data analysis

Raw RNA-sequencing data was quality assessed using FastQC[22] (v0.11.3) and aligned to the human genome (hg19) using HISAT2[23] (v2.0.4). Four samples from different individuals (two nasal, two tracheal) failed quality control checks, and were excluded from downstream analyses. Gene-level quantitation of aligned reads was performed using the `summarizeOverlaps()` function from the BioConductor *GenomicAlignments* package[24], and post-alignment quality control using SAMStat[25] (v1.5.2). Differentially expressed genes were identified between tracheal and nasal epithelial samples using the Bioconductor package *limma-voom*[26], with paired samples taken into account through the `duplicateCorrelation()` function. Differentially expressed genes were identified between study groups using the Bioconductor package DESeq2[27], with model adjustments for age and sex. P-values for both analyses were adjusted for multiple testing using the Benjamini-Hochberg method, and those <0.05 were considered significant. The weighted gene co-expression network analysis (WGCNA)[28] algorithm was used to construct signed networks for nasal and tracheal epithelial cells separately, based on gene expression from all study participants. Biological characterisation of the differentially expressed genes or consensus network modules was performed using InnateDB[29] to identify significantly over-represented pathways. For comparison to the skin dataset, raw fastq sequencing files were downloaded from the Gene Expression Omnibus[21] (GSE54456) and processed through the same pipeline as our airway epithelial samples, for optimal comparison between tissues. The skin dataset was filtered and normalised using identical methods as for the airway epithelial samples, as described in text.

Statistical methods

Statistical analyses of demographic data were performed using Prism (v7; GraphPad Software). We used the Kruskal-Wallis test with Dunn's multiple comparisons test for continuous variables (data for each group were checked using the D'Agostino & Pearson normality test) and Chi-square tests for categorical variables. All other statistics obtained outside of InnateDB or Ingenuity Pathway Analysis were performed in the R environment (v3.3.3). Correlation of average gene expression between tracheal and nasal epithelium was using the Kendall rank correlation coefficient for non-parametric data. The `modulePreservation()` function (which calculates pair-wise module preservation statistics) was used to compare individual co-expression networks for tracheal and nasal epithelium, where results are summarised into a Z-score of overall module preservation between the networks.

RESULTS

Comparison of gene expression profiles between tracheal and nasal epithelium, independent of wheeze or atopy.

Demographic data for all study participants is presented in Table 1, and a summary of all analyses performed is presented in Figure E1. RNA-sequencing resulted in an average of 20 million mapped reads per sample (range 13.7–25 million). To initially assess the composition of tracheal and nasal epithelium across all samples, we compared the overlap of expressed genes between the two sites, using a count per million (CPM) threshold corresponding to a count of at least ten in all nasal or tracheal samples. This is a standard filtering strategy for removing low or non-expressed genes[30]. We identified 16,645 and 16,720 genes expressed in tracheal and nasal epithelium respectively (independent of wheeze or atopy) of which 15,864 (~91%) were expressed in both sites, while the remainder were more frequently expressed in either site (Figure 1A). Overall, gene expression was highly correlated between tracheal and nasal epithelium (Kendall's tau = 0.82, p-value < 2.2E-16) (Figure 1B). Next, to confirm if our gene expression profiles were specific to the airway, or common to other types of epithelium, we compared our dataset to that derived from healthy skin punch biopsies[31] (n=82). Using the same filtering strategy as for the airway samples, we identified 15,784 genes expressed in the skin epithelium, of which 74–75% were also expressed in the airway epithelium (Figure E2). In general, weaker correlations were observed between gene expression profiles from the skin and each airway tissue (Kendall's tau = 0.46 and 0.50 for correlations with tracheal and nasal samples, respectively; Figure E2). In addition, to more thoroughly investigate the differences between airway sites, we tested for differentially expressed genes between tracheal and nasal epithelial samples independent of wheeze/atopy, and identified 2,467 genes with an absolute fold-change difference in expression >2 between tissues (Table E1). We then ranked these genes by highest absolute fold-change, to

characterise the most divergent genes. Genes with the greatest dichotomy in expression that were most highly expressed in tracheal epithelium included *MARCO* (involved in innate immune responses), *FABP4* (functions in lung inflammation/angiogenesis), the surfactant protein-coding gene *SFTA3* and *NKX2-1* (transcription factor involved in lung development). Those most highly expressed in nasal epithelium included a range of transcription factors (*FOXG1*, *PAX7*, *OTX2*, *SIX3*, *SP8*, *PAX3*) alongside the keratin encoding gene *KRT24*.

We next constructed weighted co-expression gene networks for tracheal and nasal epithelium in parallel (across all samples), as a holistic approach to understanding the functional organisation of the transcriptome at each site[32]. This analysis works to identify clusters (or modules) of genes with similar patterns of gene expression across each tissue, and provides biological insight based on the principal that genes with highly correlated patterns of expression are likely to participate in the same biological process. Here, we identified 17 network modules within tracheal epithelium, and 8 modules within the nasal epithelium. We next identified significantly over-represented biological pathways and gene ontology terms for each module in the tracheal and nasal network (Tables E2 and E3 in this article's Online Repository at www.jacionline.org). A comparison of these two site-specific networks revealed a significant overlap in structure, i.e. each nasal network module corresponded to one or more tracheal network modules (Figure 2A). This was confirmed through additional module preservation statistics which test how well the modules from one network, are preserved in another. Each network module from the nasal network was significantly preserved in the tracheal network (Table 2) and vice versa (Table 3); a Z-score >10 indicates strong evidence of module preservation, and the average across networks was 36.8 (range 14–67). Given the similarities between the tracheal and nasal networks, we next built a consensus network, where the resulting network modules represent robust gene co-expression patterns

common to both tracheal and nasal epithelium. The consensus network contained 16 modules of co-expressed genes (Figure 2B), and each module of the consensus network significantly overlapped with at least one module from both tracheal and nasal epithelial networks (Figure E3 in this article's Online Repository at www.jacionline.org). Most of the consensus network modules were associated with generalised cellular processes (ribosomal function, metabolism, gene expression, mitochondrial activity), while two modules were associated with immune responses, one being specific for interferon mediated anti-viral responses (Table E4 in this article's Online Repository at www.jacionline.org). To assess the relative expression of each consensus network module in each tissue and within each study group, the median fold-change difference in expression of the consensus module genes was measured between tracheal and nasal samples, and was less than one across all phenotypes (Figure 2C). This suggests a modest difference in the relative expression of consensus modules between tissues, independent of atopy/wheeze (Figure E4 this article's Online Repository at www.jacionline.org presents the results for each study group individually). Of note, modules associated with cilium assembly and cell cycle activity (yellow and blue) were more highly expressed in the tracheal epithelium, while nasal epithelium was associated with increased expression of modules involving growth and metabolism (red and brown). Taken together, these results demonstrate overall conservation of the underlying transcriptional networks between tracheal and nasal epithelium, while specific gene expression differences relate to site-specific functions.

Gene expression changes associated with wheeze and atopy occur in both nasal and tracheal epithelium

To identify gene expression changes associated with atopy and or wheeze within the epithelium, differential gene expression analyses were performed contrasting each group

against healthy controls. Specifically, these comparisons inform on the gene expression changes related to the combined effect of wheeze and atopy (wheeze with atopy vs healthy controls), atopy only (atopy vs healthy controls) or wheeze only (wheeze vs healthy controls). As an alternative approach to identifying gene signatures attributed to wheeze only, we also compared atopics with or without wheeze. The results from each of the four comparisons across both tracheal and nasal epithelium are summarised in Table 4 (Tables E5–E12 in this article's Online Repository at www.jacionline.org). Comparisons related to the combined effect of atopy and wheeze yielded the largest gene expression differences in both tissues, while overall, disease associated changes were more prominent in the tracheal epithelium (greater numbers of differentially expressed genes compared to nasal epithelium). Importantly, we observed a partial overlap between the lists of genes generated from each contrast across tissues, which overall showed strong agreement in terms of fold-change values (Tables E13–E16 and Figure E5 in this article's Online Repository at www.jacionline.org). The greatest overlap was evident for contrasts between wheeze with atopy and healthy controls; 49 genes were differentially expressed in both the lower and upper airway epithelium, including several well characterised asthma-associated genes (*IL1RL1*, *PTGS1*, *CCL26* and *POSTN*).

Lower airway gene-signatures in atopic wheezers can distinguish between groups in upper airway samples

We next sought to obtain a more detailed understanding of the molecular mechanisms that characterise the wheeze-associated signatures in the upper and lower airways through WGCNA, which provides a holistic view of the functional organisation of the gene expression program. Separate networks were constructed for the tracheal and nasal epithelial samples based on genes with significant variation in expression across samples (to capture genes with

intrinsic biological variation between phenotypes). The inclusion of a variance-based filter contrasts with the previous network analysis, which was performed on a transcriptome-wide scale. This analysis defined nine modules (or clusters) of co-expressed genes within the tracheal epithelium (Figure 3A), and six modules within the nasal epithelium (Figure E6A in this article's Online Repository at www.jacionline.org). To determine which network module(s) were perturbed in relation to wheeze or atopy, each module was checked for enrichment of differentially expressed genes across all four contrasts. Module E within the tracheal epithelial network was highly enriched for genes associated with wheeze with atopy, while no other network module (in either tissue) was associated with a disease phenotype (Figure 3B and Figure E6B-E in this article's Online Repository at www.jacionline.org). Herein, module E is referred to as the tracheal wheeze with atopy module. To test whether differences in these genes were specific to the lower airway, or were also present in the upper airway, we performed semi-supervised clustering analysis using the 92 genes. Heatmap visualisation of the 92 genes within this module indicated that overall i) the genes were more highly expressed in individuals with wheezing with atopy or atopy alone and ii) this expression pattern and sample clustering was consistent within nasal epithelium (Figure 3C). Importantly, Principal Component Analysis based on the 92-genes within the tracheal wheeze with atopy module, could equally distinguish between groups in nasal epithelium (Figure 3D). Of note, in both upper and lower airway samples, subjects with wheeze and atopy or atopy alone tended to cluster together, while subjects with wheeze alone clustered more closely with the healthy controls. This is consistent with the fact that none of the genes within the tracheal wheeze with atopy module were significantly differentially expressed in relation to wheeze alone in either tracheal or nasal epithelium (Table E17 in this article's Online Repository at www.jacionline.org).

Children with atopy exhibit similar gene expression profiles across nasal and tracheal epithelium

To capture the most robust changes associated with wheezers with atopy, the twenty most significantly differentially expressed genes between children with wheeze and atopy and healthy controls, and the twenty genes with the greatest absolute fold-change difference in expression were identified in tracheal and nasal epithelium i.e. genes were ranked once by adjusted p-value, then again by absolute fold-change for each tissue, resulting in four lists. Five genes (*CST1*, *CST2*, *CST4*, *CLCA1*, *POSTN*) were consistently ranked in the top twenty genes across all four lists, all of which were significantly up-regulated in children with atopy (Figure 4). In addition, a first-order protein-protein interaction network containing the tracheal wheeze with atopy module genes was created to identify the most highly interconnected genes (hubs; representing key molecules within a network) (Figure E7 in this article's Online Repository at www.jacionline.org). Again, the atopic groups exhibited increased expression for several hub genes (*IL1R1*, *NTRK1*, *KRT1*) across both nasal and tracheal airway epithelium (Figure 4).

DISCUSSION

The upper and lower respiratory tracts form a continuum allowing the passage of air into and out of the lungs. Despite epidemiological and pathophysiological evidence to suggest a unified airway, there are also significant differences between these two regions. Although a unified airway has been inferred, there is conjecture on whether this is in fact the case due to a lack of direct evidence. The current study is the largest of its kind conducted to directly test this hypothesis by undertaking a global transcriptomic analysis on freshly isolated nasal and tracheal epithelial cells from children, with and without atopy and/or wheeze. Our observations support the hypothesis of unified airway with 91% homology in the types of genes being expressed in the upper and lower airway. Furthermore, network analysis illustrated conservation of co-expressed (and functionally enriched) gene modules between the upper and lower airway which were associated with generalised cellular processes including; ribosomal function, metabolism, and mitochondrial activity. Furthermore, we identified gene expression changes associated with wheeze and/ or atopy (*IL1RL1, PTGS1, CCL26, POSTN, CST1, CST2, CST4, CLCA1, NTRK1*) that are inherent to both the upper and lower airway epithelium. Collectively, the findings support the view of a unified airway at the transcriptomic level and provide rationale for future exploration into functional relationships between these two regions.

The few studies that have used transcriptomics to compare nasal and tracheal epithelial cells have done so under specific airway disease settings including asthma, cystic fibrosis (CF) and Chronic Obstructive Pulmonary Disorder (COPD), have utilised relatively small sample sizes (<15), and compared their data with publicly available transcriptomic data sets[33, 34]. In contrast, our study performed transcriptomic analysis on a paediatric cohort of 63 children with and without wheeze and/or atopy, using predominantly matched nasal and tracheal

samples. This study was also designed and conducted with high stringency with only those RNA samples meeting the set quality (ratio >1.80), integrity (RIN>7.8) and quantity (>800ng) standards subsequently sequenced. Procedures were also harmonized for nasal and airway sampling as well as the processing and extensive training at the initial phase of the study ensured that the method and location of sampling as well as the subsequent processing was standardized. We have also minimized sampling and technique confounders to the study by employing only 1 physician to obtain all samples using the one nylon cytology brush. As a result, we feel this study has a unique robustness that has enabled detailed and meaningful analysis of its findings.

Not surprisingly, we identified a family of keratin genes (*KRT23, KRT33A, KRT34, KRT36, KRT37, KRT71*) that were not differentially expressed between sites. Being archetypal lineage markers, broad commonality between the nasal and tracheal epithelium of these genes support the concept of a common cell lineage and unified airway. Furthermore, observed similarities between the nasal and tracheal epithelium in terms of modules/clusters of co-expressed genes also highlights the globally conserved biological functions across both types of epithelium irrespective of disease phenotype, although the higher number of modules initially defined in tracheal epithelium may reflect a higher level of functional complexity in the lower airway. For example, several tight junction gene expression levels (*CDH10, CDH15, CDH19, CDH9, PCDH11X, PCDH8*) were similarly expressed in both the upper and lower airway and suggest a commonality in barrier function maintenance.

Further exploration of the genes more highly expressed in the upper airway found that in addition to olfactory genes (*ORFE129P, ORF7E158P, ORF7E91P, ORF7E99P*) and genes related to epithelial lineage (*KRT16, KRT16P3, KRT18P32, KRT8P5, KRT2, KRT24,*

KRT8P1, KRT8P38, KRTP5-9), were genes from the uridine 5'-diphospho-glucuronosyltransferase (UDP) family (*UGT1A10, UGT1A7, UGT1A8, UGT1A2, UGT1A3, B3GALT1, B3GNT6, GALNT13, GALNT16, GALNT4).* These genes have been found to be expressed in the lung epithelium [35, 36] in addition to hepatic tissue and collectively act to remove toxins, endogenous substances and foreign chemicals, which align to the primary function of the upper airway [37, 38]. Tracheal epithelial cells on the other hand were found to more highly express tubulin genes (*TUBAL3, TUBB8P7*) whose purpose is to direct airway ciliary function [39 40]. This is corroborated further by the observation that network modules related to cilium assembly/activity were more highly expressed in the lower airway. Also prominent in the lower airway were genes related to calcium-activated chloride channels (*ANO3, ANO4, ANO5, BEST3, CADPS, CACNG4, CACNA1B, CACNA1E, CACNA1H*) which regulate and maintain the air surface liquid (ASL) layer at the luminal surface of epithelia, which is critical for effective mucociliary clearance [41 42]. We then broadened our comparisons to include the skin in order to determine whether the predominantly unified signature extended beyond the airway into other epithelial tissues. We identified a lower level of similarity (75%) in the genes expressed in the skin and the nasal/tracheal epithelium compared to the similarity observed between nasal and tracheal epithelium (95%), which was corroborated by weaker correlations observed in the gene expression profiles between the two sites. Collectively, these results suggest that while the upper and lower airway exhibit conservation of gene expression for most genes, differential regulation of gene expression affecting a small subset of genes is evident between the two sites, reflecting site-specific functions. Furthermore, and probably not surprisingly, our data show that even though there is overlap in the gene signature patterns between respiratory and non-respiratory epithelium, there are unique features in the airway that do not extend to other types of epithelium.

We believe this is the largest study of its kind to compare transcriptional profiles of airway epithelium between children that exhibit wheeze and atopy using paired nasal and tracheal primary airway epithelial samples. Our finding of a conserved signature across both epithelial tissues in children with wheeze and atopy supports other investigations observing similar transcriptional profiles between nasal and tracheal tissues in the context of lower airway pathologies including; childhood asthma[43, 44], COPD[45] and those induced through cigarette-smoke[46, 47].

Interestingly, several identified genes in our wheeze with atopy group have been also associated with lower airway pathology, namely asthma. Considering the processes that lead to asthma development start very early in life and persistent wheezing combined with allergies is a strong-risk factor for its development, overlaps in the expression of particular genes may be expected. Of relevance to this study, are the five genes representing the most robust changes in expression associated with wheeze and atopy across both nasal and tracheal epithelium (*CST1*, *CST2*, *CST4*, *CLCA1* and *POSTN*). *CST1*, *CST2* and *CST4* encode for the closely related cystatin-S proteins[48] and have previously been reported as significantly upregulated in bronchial epithelium of mild to moderate asthmatics[49], and in nasal epithelium from individuals with allergic rhinitis (*CST1* only[50]). Cystatins are a large family of peptides that are potent inhibitors of cysteine proteinases and are often referred to as "salivary cystatins" due to their recognised role as defence proteins within in saliva[51], even though cystatin-S proteins are also expressed in both bronchial and nasal lavage fluids[52]. Another identified gene, *CLCA1* (chloride channel accessory 1) belongs to a family of calcium-sensitive chloride conductance proteins and as such, is a critical physiological regulator of chloride transport and mucin expression; *CLCA1* is highly up-regulated in response to IL-13 and can drive mucus production in human airway epithelial cells[53]. The role of *CLCA1* in

hypersecretory respiratory diseases including asthma, COPD and cystic fibrosis is well documented[54, 55]. Finally, POSTN encodes for periostin; a secreted extracellular matrix protein that is a well-investigated marker of airway remodelling[56]. POSTN expression is induced downstream of IL-4/IL-13 signalling in human airway epithelial cells[57] and plays a role in wound healing and pro-fibrogenic processes[58]. Importantly, over-expression of POSTN delays wound healing in mice[59] and promotes airway hyperresponsiveness and inflammation[60]. Collectively, the fact that the conserved epithelial-signature associated with wheeze and atopy identified in this study contains many genes previously identified as relevant to asthma pathogenesis, highlights the clinical relevance of our findings.

Several other genes within our conserved epithelial-signature are hallmarks of Th2-driven inflammation often observed in asthma or are genes in which genetic variation is significantly associated with asthma risk (*CCL26*, , *IL1R1 and NTRK1*)[56,61-63]. Together, these results suggest that there is a common mechanism underlying the intrinsic changes that extend between the upper and lower airway epithelium in atopic asthmatic children, and while the differences may only affect a subset of genes (49 in total), they represent functionally important changes that are directly relevant to lower airway pathology.

Our study is not without limitations. Firstly, our study samples were collected from a community cohort, and although subgroups could be identified, better definition of these using strong accompanying criterion would strengthen the ability to compare results between these groups. Secondly, our sampling methodology obtained epithelial cells from the central airway, namely just above the carina. We were unable to sample deeper in the lung and thus our interpretations of a 'unified airway' are supportive of but limited to the central airway. Future studies may explore testing the hypothesis and comparing nasal and peripheral

bronchiolar epithelium. Thirdly, although our sampling methodology yields a >95% pure epithelial cell population, we acknowledge that our samples may not be entirely pure. Assessments of collected samples have identified a small population of macrophages (1-5% of sample) as well as a heterogeneous epithelial cell population consisting of both terminally differentiated epithelial cells as indicated by the presence of cilia and non-ciliated basal epithelial cells[19]. Our routine sample processing methodology positively selects out the macrophages[18, 19] and interrogation of resulting cell populations have confirmed that samples are not contaminated by other cell types, including; mesenchymal cells, macrophages, dendritic cells and endothelial cells[18]. Although these have been excluded we recognize that we may potentially have other cells types that would still be detected by RNA-seq. This is particularly so for immune cells including lymphocytes and monocytes which are known to have well characterized pathways/gene ontologies[64, 65]. In fact, modules identified as part of this study included pathways involved in T cell signaling and B cell signaling. Although beyond the scope of the current manuscript, future experiments could utilise single-cell sequencing to delineate the exact transcriptomic signature for each cell type in the airway epithelium[66, 67]. Finally, atopy was determined in this study using a RAST test. This was justified since it was directly measured at time of recruitment and additional information such as eczema and hayfever was only parent reported. Future studies would be strengthened if this information was confirmed as current manifestations and combined with RAST results to further define the atopy group.

Although data generated from this study indicates primary homology between the upper and lower airway, the importance of the unified airway hypothesis lies in its approach to therapy. The hypothesis assumes that commonalities between the nasal and tracheal mucosa are functionally complementary and thus can be treated in a unified fashion. Our observations

suggest that this the case, with biological conservation evident. Considering, since clinical evidence suggests that treating upper airway diseases such as rhinitis benefits the lower airway[68-70], there is inference that certain functional complementarity between the nose and the lower airway may facilitate a unified approach to treatment.

Conclusions and future directions

Collectively, this study, the largest of its kind, has utilized transcriptomics to comprehensively and systematically compare upper and lower airway epithelial cells. Data generated identified 91% homology with the remainder being uniquely expressed to each region and associated with site-specific roles. There exists enormous interest as to whether nasal cells can act as an informative surrogate for the lower airway. Our findings suggest that this may be the case but dependent on whether both sites are affected to a similar extent and if both translate to similar functional outcome. With significant clinical interest also in biomarker discovery, and the use of nasal cells as an easily accessible surrogate for lower airway tissues, the conserved epithelial-signature identified in this study holds promise for development into a screening tool. While further work is required to determine the predictive value of these genes as biomarkers for chronic airway diseases, this unique data set provides a valuable resource for future meta-analyses or integrated -omic studies.

ACKNOWLEDGEMENTS: The authors thank the participants and their families in the Western Australian Epithelial Research Program (WAERP) for their contribution to this study. The members of WAERP include are: Anthony Kicic, Stephen M Stick, Darryl A Knight, Elizabeth Kicic-Starcevich, Luke W Garratt, Marc Padros-Goosen, Ee-Lyn Tan, Erika N Sutanto. Kevin Looi, Jessica Hillas, Thomas Iosifidis, Nicole C Shaw, Samuel T Montgomery, Kak-Ming Ling, Kelly M Martinovich, Francis J Lannigan, Ricardo Bergesio,

Bernard Lee, Shyan Vijaya-Sekeran, Paul Swan, Mairead Heaney, Ian Forsyth, Tobias Schoep, Alexander Larcombe, Monica Hunter, Kate McGee, Nyssa Millington. The Australian Respiratory Epithelium Consortium (AusREC) acknowledges the following members: Anthony Kicic, Stephen M Stick, Elizabeth Kicic-Starcevich, Luke W Garratt, Erika N Sutanto, Kevin Looi, Jessica Hillas, Thomas Iosifidis, Nicole C Shaw, Samuel T Montgomery, Kak-Ming Ling, Kelly M Martinovich, Matthew W-P Poh, Daniel R Laucirica, Craig Schofield, Samantha McLean, Katherine Landwehr, Emma de Jong, Nigel Farrow, Eugene Roscioli, David Parsons, Darryl A Knight, Christopher Grainge, Andrew T Reid, Su-Ling Loo, and Punnam C Veerati.

REFERENCES

1. Meena RS, Meena D, Aseri Y, Singh BK, Verma PC. Chronic Rhino-Sinusitis and Asthma: Concept of Unified Airway Disease (UAD) and its Impact in Otolaryngology. Indian J Otolaryngol Head Neck Surg 2013;65:338-342.

2. Krouse JH. The Unified Airway—Conceptual Framework. Otolaryngol Clin North Am 2008;41:257-266.

3. Mygind N. Rhinitis and asthma: similarities and differences., ed. P.U. Mygind N, Pipkorn U, Dahl R. 1990, Copenhagen: Munksgaard.

4. Giavina-Bianchi P, Aun MV, Takejima P, Kalil J, Agondi RC. United airway disease: current perspectives. J Asthma Allergy 2016;9:93-100.

5. Bachert C, Vignola AM, Gevaert P, Leynaert B, Van Cauwenberge P, Bousquet J. Allergic rhinitis, rhinosinusitis, and asthma: one airway disease. Immunol Allergy Clin North Am 2004;24:19-43.

6. Fontanari P, Zattara-Hartmann MC, Burnet H, Jammes Y. Nasal eupnoeic inhalation of cold, dry air increases airway resistance in asthmatic patients. Eur Respir J 1997; 10: 2250-2254.

7. Jörres R, Nowak D, Magnussen H. The effect of ozone exposure on allergen responsiveness in subjects with asthma or rhinitis. Am J Respir Crit Care Med 1996;153:56-64.

8. Michel FB, Marty JP, Quet L, Lour P. Penetration of Inhaled Pollen into the Respiratory Tract 1, 2. Am Rev Respir Dis 1977;115:609-616.

9. Platts-Mills TA. The role of allergens in allergic airway disease. J Allergy Clin Immunol 1998;101:S364-S366.

10. Chanez P, Vignola AM, Vic P, Guddo F, Bonsignore G, Godard P et al. Comparison between nasal and bronchial inflammation in asthmatic and control subjects. Am J Respir Crit Care Med 1999;159:588-595.

11. Gaga M, Lambrou P, Papageorgiou N, Koulouris NG Kosmas E, Fragakis S et al. Eosinophils are a feature of upper and lower airway pathology in non-atopic asthma, irrespective of the presence of rhinitis. Clin Exp Allergy 2000;30:663-669.

12. Braunstahl GJ, Overbeek SE, Fokkens WJ, Kleinjaqn A, McEuen AR, Walls AF et al. Segmental bronchoprovocation in allergic rhinitis patients affects mast cell and basophil numbers in nasal and bronchial mucosa. Am J Respir Crit Care Med 2001;164:858-865.

13. Braunstahl GJ, Overbeek SE, Kleinjan A, Prins JB, Hoogsteden HC, Fokkens WJ. Nasal allergen provocation induces adhesion molecule expression and tissue eosinophilia in upper and lower airways. J Allergy Clin Immunol 2001;107:469-476.

14. Bourdin A, Gras D, Vachier I Chanez P. Upper airway· 1: Allergic rhinitis and asthma: united disease through epithelial cells. Thorax 2009;64:999-1004.

15. Togias A. Rhinitis and asthma: evidence for respiratory system integration. J Allergy Clin Immunol 2003;111:1171-1183.

16. Asher MI, Keil U, Anderson HR, Beasley R, Crane J, Martinez F et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. Eur Respir J 1995;8:483-491.

17. Ferris BG. Epidemiology Standardization Project (American Thoracic Society). Am Rev Respir Dis 1978; 118:1-120.

18. Kicic A, Sutanto EN, Stevens PT, Knight DA, Stick SM. Intrinsic biochemical and functional differences in bronchial epithelial cells of children with asthma. Am J Respir Crit Care Med 2006;174:1110-1118.

19. Lane C, Burgess S, Kicic A, Knight DA, Stick SM. The use of non-bronchoscopic brushings to study the paediatric airway. Respir Res 2005;6:53-63.

20. Lane C, Knight DA, Burgess S, Franklin P, Horak F, Legg J *et a*l. Epithelial inducible nitric oxide synthase activity is the major determinant of nitric oxide concentration in exhaled breath. Thorax 2004;59:757-760.

21. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-210.

22. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. [cited 2017; Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

23. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357-360.

24. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R *et al*. Software for Computing and Annotating Genomic Ranges. PLoS Compu Biol 2013;9:e1003118.

25. Lassmann T, Hayashizaki Y, Daub CO. SAMStat: monitoring biases in next generation sequencing data. Bioinformatics 2011;27:130-131.

26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

27. Love MI, Huber W, Anders S. Moderated estimatation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559-559.

29. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic Acids Res 2013;41:D1228-D1233.

30. Chen Y, Lun AT, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res 2016;5:1438.

31. Li B, Tsoi LC, Swindell WR, Gudjonsson JE, Tejasvi T, Johnston A et al. Transciptome analysis of psoriasis in a large case-control sample: RNA-seq provides insights into diseas mechanisms. J. Invest. Dermatol 2014;134:1828-1838.

32. Jones AC, Bosco A. Using Network Analysis to Understand Severe Asthma Phenotypes. Am J Respir Crit Care Med 2017;195:1409-1411.

33. Clarke LA, Sousa L, Barreto C, Amaral MD. Changes in transcriptome of native nasal epithelium expressing F508del-CFTR and intersecting data from comparable studies. Respir Res 2013;14:e38.

34. Ogilvie V, Passmore M, Hyndman L, Jones L, Stevenson B, Wilson A et al. Differential global gene expression in cystic fibrosis nasal and bronchial epithelium. Genomics 2011;98:327-336.

35. Court MH. Interindividual variability in hepatic drug glucuronidation: studies into the role of age, sex, enzyme inducers, and genetic polymorphism using the human liver bank as a model system. Drug Metab Rev, 2010;42:209-224.

36. Ohno S, Nakajin S. Determination of mRNA expression of human UDP-glucuronosyltransferases and application for localization in various human tissues by real-time reverse transcriptase-polymerase chain reaction. Drug Metab Dispos, 2009;37:32-40.

37. Heydel JM, Holsztynbska EJ, Legendre A, Thiebaud N, Artur Y, Le Bon AM. UDP-glucuronosyltransferases (UGTs) in neuro-olfactory tissues: expression, regulation, and function. Drug Metab Rev, 2010;42:74-97.

38. Ouzzine M, Gulberti S, Ramalanjaona N, Magdalou Jm Fournel-Gigleux S. The UDP-glucuronosyltransferases of the blood-brain barrier: their role in drug metabolism and detoxication. Front Cell Neurosci, 2014;8:e349.

39. Ross, AJ, Dailey LA, Brighton LE, Devlin RB. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. Am J Respir Cell Mol Biol, 2007;37:169-185.

40. Jain R, Pan, J, Driscoll JA, Wisner JW, Huang T Gunsten SP et al. Temporal relationship between primary and motile ciliogenesis in airway epithelial cells. Am J Respir Cell Mol Biol, 2010;43:731-739.

41. Wagner JA, Cozens AL, Schulman H, Gruenert DC, Stryer L, Gardner P. Activation of chloride channels in normal and cystic fibrosis airway epithelial cells by multifunctional calcium/calmodulin-dependent protein kinase. Nature, 1991;349:793-796.

42. Huang F, Zhang H, Wu M, Yang H, Kudo M, Peters CJ et al. Calcium-activated chloride channel TMEM16A modulates mucin secretion and airway smooth muscle contraction. Proc Natl Acad Sci USA, 2012;109:16354-16359.

43. Guajardo JR, Schleifer KW, Daines MO, Ruddy RM, Aronow BJ, Wills-Karp M et al.. Altered gene expression profiles in nasal respiratory epithelium reflect stable versus acute childhood asthma. J Allergy Clin Immunol 2005;115:243-251.

44. Poole A, Urbanek C, Eng C, Schageman J, Jacobson S, O'Connor BP, et al. Dissecting Childhood Asthma with Nasal Transcriptomics Distinguishes Subphenotypes of Disease. J Allergy Clin Immunol 2014;133:670-678.

45. Boudewijn IM, Faiz A, Steiling K, van der Wiel E, Telenga ED, Hoonhorst SJ, et al.. Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression. Respir Res 2017;18:213.

46. Zhang X, Sebastiani P, Liu G, Schembri F, Zhang X, Dumas YM, et al. Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. Physiol Genomics 2010;41:1-8.

47. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K, et al. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. BMC Genomics 2008;9:259.

48. Dickinson D. Salivary (SD-type) cystatins: over one billion years in the making—But to what purpose? Crit Rev Oral Biol Med 2002;13:485-508.

49. Singhania A, Wallington JC, Smith CG, Horowitz D, Staples KJ, Howarth PH, et al.. Multitissue Transcriptomics Delineates the Diversity of Airway T Cell Functions in Asthma. Am J Respir Cell Mol Biol 2018;58:261-270.

50. Imoto Y, Tokunaga T, Matsumoto Y, Hamada Y, Ono M, Yamada T, et al. I. Cystatin SN Upregulation in Patients with Seasonal Allergic Rhinitis. PLoS One 2013;8:e67057.

51. Fábián TK, Hermann P, Beck A, Fejérdy P, Fábián G. Salivary Defense Proteins: Their Network and Role in Innate and Acquired Oral Immunity. Int J Mol Sci 2012;13:4295-4320.

52. Lindahl M, Ståhlbom B, Tagesson C. Newly identified proteins in human nasal and bronchoalveolar lavage fluids: potential biomedical and clinical applications. Electrophoresis 1999;20:3670-3676.

53. Alevy YG, Patel AC, Romero AG, Patel DA, Tucker J, Roswit WT,et al. IL-13–induced airway mucus production is attenuated by MAPK13 inhibition. Journal Clin Invest 2012;122:4555-4568.

54. Brett TJ. CLCA1 and TMEM16A: The link towards a potential cure for airway diseases. Expert Rev Respir Med 2015;9:503-506.

55. Winpenny JP, Marsey LL, Sexton DW. The CLCA gene family: putative therapeutic target for respiratory diseases. Inflammation & Allergy-Drug Targets (Formerly Current Drug Targets-Inflammation & Allergy) 2009;8:146-160.

56. Zissler UM, Esser-von Bieren J, Jakwerth CA, Chaker AM, Schmidt-Weber CB. Current and future biomarkers in allergic asthma. Allergy 2016;71:475-494.

57. Yuyama N, Davies DE, Akaiwa M, Matsui K, Hamasaki Y, Suminami Y, et al. Analysis of novel disease-related genes in bronchial asthma. Cytokine 2002;19:287-96.

58. Walker JT, McLeod K, Kim S, Conway SJ, Hamilton DW. Periostin as a Multifunctional Modulator of the Wound Healing Response. Cell and tissue research 2016;365:453-465.

59. Nunomura S, Nanri Y, Ogawa M, Arima K, Mitamura Y, Yoshihara T et al. Constitutive overexpression of periostin delays wound healing in mouse skin. Wound Repair and Regeneration 2018;26:6-15.

60. Bentley JK, Chen Q, Hong JY, Popova AP, Lei J, Moore BB et al. Periostin is required for maximal airways inflammation and hyperresponsiveness in mice. J Allergy Clin Immunol. 2014;134:1433-1442.

61. Bhakta NR, Solberg OD, Nguyen CP, Nguyen CN, Arron JR, Fahy JV, et al. A qPCR-based metric of Th2 airway inflammation in asthma. Clin Translational Allergy 2013;3:24-24.

62. Traister RS, Uvalle CE, Hawkins GA, Meyers DA, Bleecker ER, Wenzel SE. Phenotypic and genotypic association of epithelial IL1RL1 to human T(H)2-like asthma. J Allergy and Clin Immunol 2015;135:92-99.e10.

63. Szczepankiewicz A, Rachel M, Sobkowiak P, Kycler Z, Wojsyk-Banaszak I, Schöneich N. Neurotrophin serum concentrations and polymorphisms of neurotrophins and their receptors in children with asthma. Respir Med 2013;107:30-36.

64. Reynolds LM, Ding J, Taylor JR, Lohman K, Soranzo N, de la Fuente A et al. Transcriptomic profiles of aging in purified human immune cells. BMC Genomics 2015;16:e333.

65. Mirsafian H, Ripen AM, Leong WM, Manaharan T, Mohamad SB, Merican AF.Transcriptome landscape of human primary monocytes at different sequencing depth. Genomics 2017;109:463-470.

66. Baslan T, Hicks J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. Nat Rev Cancer 2017;17:557-569.

67. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell 2015;58:610-620.

68. Agondi RC, Machado ML, Kalil J, Giavina-Bianchi P.Intranasal corticosteroid administration reduces nonspecific bronchial hyperresponsiveness and improves asthma symptoms. J Asthma 2008;45:754-757.

69. Lohia S, Schlosser RJ, Soler ZM. Impact of intranasal corticosteroids on asthma outcomes in allergic rhinitis: a meta-analysis. Allergy 2013;68:569-579.

70. Crystal-Peters J, Neslusan C, Crown WH, Torress A. Treating allergic rhinitis in patients with comorbid asthma: the risk of asthma-related hospitalizations and emergency department visits. J Allergy Clin Immunol 2002;109:57-62.

FIGURE LEGENDS

Figure 1. Conserved transcriptional profiles between tracheal and nasal epithelium. (A) Venn diagram comparing genes with evidence of expression in all tracheal, and all nasal epithelial samples, independent of asthma or atopy. **(B)** Scatter plot of the average gene counts per million (CPM) reads, for all expressed genes between tracheal and nasal epithelial samples (across all groups). Colours represent genes identified as more frequently expressed in tracheal or nasal epithelium (green and purple, respectively), or common to both sites (grey).

Figure 2. Identification of a consensus gene expression network between tracheal and nasal epithelium. (A) WGCNA was used to identify modules of co-expressed genes within the upper or lower airway. Cluster dendrograms for the tracheal and nasal networks containing 17 and 8 modules respectively, are shown alongside the correspondence matrix between the two networks. Here, numbers within the table represent the number of genes that overlap between the modules being compared, where a stronger colour indicates a more significant overlap based on a Fisher's exact test p-value. **(B)** The consensus network containing 16 modules that are shared across the tracheal and nasal networks. Each network module is represented by a different colour. **(C)** The relative expression of all genes within each consensus module represented as the median $\log_2$ fold-change between tracheal and nasal epithelial cell samples, where box plots represent the results across all four study groups i.e. comparisons between tissues were performed independently for each phenotype based on atopy/wheeze.

Figure 3. Lower airway disease-signature is reflected in the upper airway. (A) Dendrogram representing the weighted correlation network analysis across tracheal epithelial samples, with each distinct module marked by a unique colour. **(B)** Each network module was tested for enrichment of differentially expressed genes (between atopic asthmatics and healthy controls within tracheal epithelium), where the horizontal dashed line represents the threshold for significance. Module E was highly enriched for differentially expressed genes. (**C**) Heatmaps visualising expression patterns for the 92 genes within module E across all nasal epithelial samples (left) and tracheal epithelial samples (right), where genes and samples were

ranked using unsupervised clustering. (**D**) Principal component analysis across all samples based on the 92 genes within module E.

Figure 4. Similar expression of key disease-associated genes within tracheal and nasal epithelium. Normalised gene counts for selected genes representing the most robust gene expression changes across tissues, or hub genes within the protein-protein interaction network for the wheeze with atopy network module. P-values were derived from DESeq2 analysis for differential gene expression (disease group relative to healthy controls) and adjusted for multiple comparisons using the Benjamini-Hochberg method. Data points represent individual samples with bars showing median values. *$p \leq 0.05$, **$p \leq 0.001$, ***$p \leq 0.0001$, ****$p \leq 0.00001$.

Table 1. Demographic information for study participants

	Healthy controls	Wheeze with atopy	Atopy only	Wheeze only	p-value
Number of participants	**16**	**15**	**16**	**16**	
Samples from participants:					
Paired TEC and NEC	13	10	13	4	
TEC only	2	1	2	3	
NEC only	1	4	1	9	
Age: median (yr) (range, yrs)	3.3 (2.8–6.8)	6.4 (4.0–18.3)	3.8 (2.8–5.5)	3.5 (2.8–7.6)	0.052
Male (%)	50	60	50	56	0.81
RAST results (% positive):					
Grass pollen	0	80	38	0	<0.0001
Animal	0	60	38	0	<0.0001
Mould	0	40	25	0	0.0035
Egg white	0	33	44	0	0.0012
House dust mite	0	53	38	0	0.0002
Past wheeze (%)	19	93	31	100	<0.0001
Current wheeze (%)	19	80	19	75	0.0005
Hay fever (%)	6	87	44	31	<0.0001
Family history of hay fever (%)	75	87	75	75	0.83
Eczema (%)	19	60	75	37	0.0084
Family history of eczema (%)	37	53	50	75	0.19
Parental asthma (%)	37	*40	37	87	0.01
Parental smoking (%)	19	13	6	44	0.051

For comparison of age across groups, the p-value was derived from Kruskal-Wallis test with Dunn's multiple comparisons test for non-parametric data (data for each group checked using the D'Agostino & Pearson normality test). Age is reported as median (inter-quartile range). For comparisons of categorical variables, p-values were derived from a Chi-square test across the four groups. All percentages are rounded to the nearest whole number. *Data was missing for one participant. TEC, tracheal epithelial cells; NEC, nasal epithelial cells.

Table 2. Nasal epithelial cell network module preservation within tracheal epithelial cell network modules

Nasal Network Module		Generalised biological function (Supplementary Table 1)	Module size	Preservation Z-score summary
1	Brown	Immune response, antigen presentation	1484	57.8
2	Green	Ribosome activity	947	56.6
3	Red	Innate immune response, interferon signalling	350	56.5
4	Blue	Cell cycle, centrosome activity	3738	49.1
5	Yellow	Adaptive immune response, B and T cell activity	1030	38.5
6	Turquoise	Metabolism, growth factor signalling	3845	15.6
7	Black	Cell cycle	185	15.2
8	Grey	No collective function (contains genes not associated with any module)	371	2.3
Total genes in network			11,950	

Table 3. Tracheal epithelial cell network module preservation within nasal epithelial cell network modules

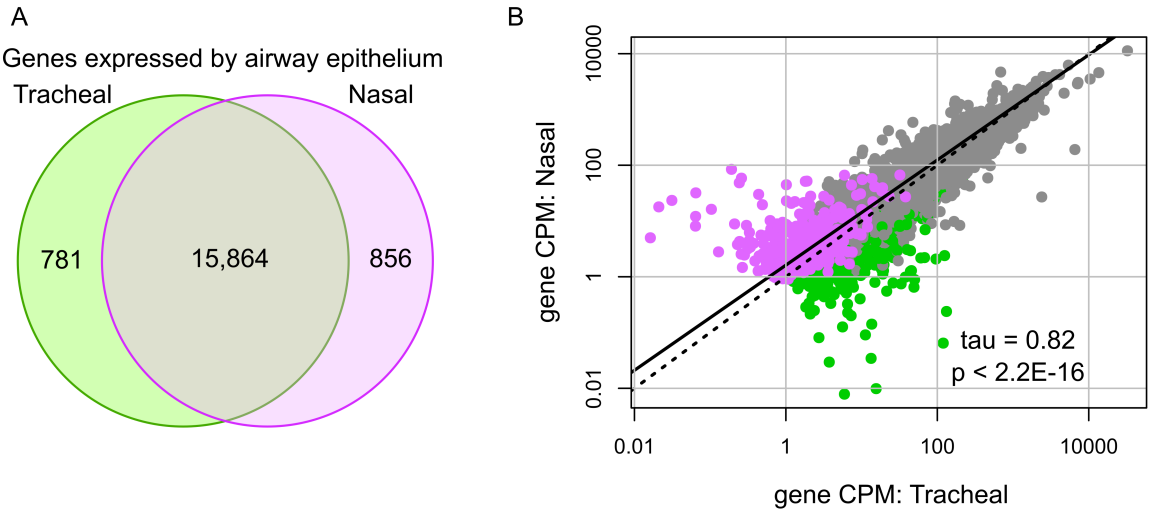
Tracheal Network Module		Generalised biological function (Supplementary Table 2)	Module size	Preservation Z-score summary
1	Cyan	Regulation of transcription	379	67.0
2	Black	Cilium/cell cycle	789	61.3
3	Magenta	Cilium activity and hedgehog signalling	610	58.6
4	Red	Ribosome activity	859	46.3

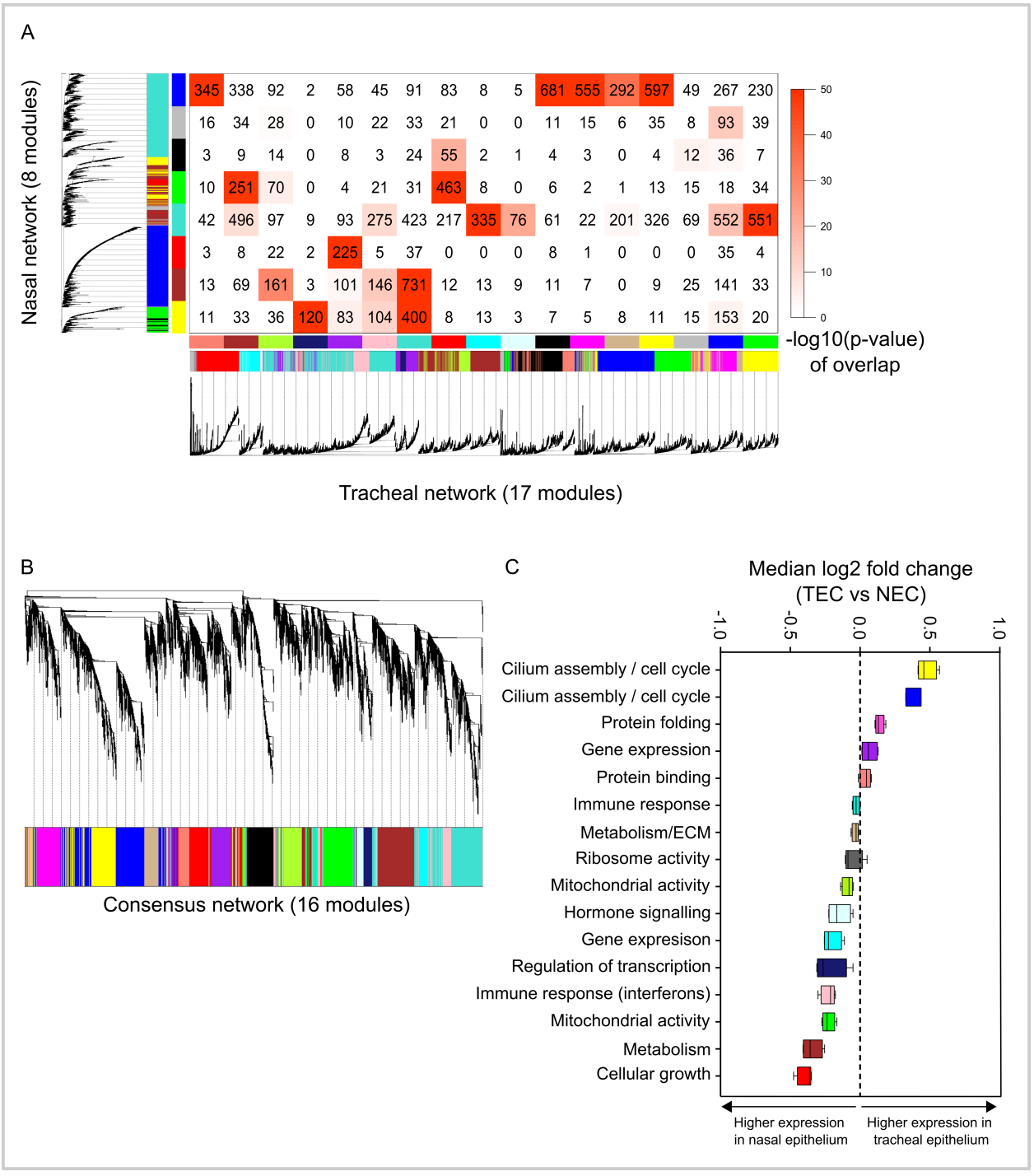
5	Turquoise	Immune response	1770	45.4
6	Midnight blue	Immune response (adaptive)	136	39.1
7	Salmon	Cell cycle/movement	443	34.1
8	Light cyan	Apoptosis	94	30.7
9	Purple	Anti-viral immune response (interferons)	582	30.4
10	Tan	NOTCH signalling/metabolism	508	27.9
11	Brown	Mitochondrial activity	1238	27.5
12	Yellow	Transcription, gene expression	995	24.1
13	Green yellow	Gene expression, proteasome activity	520	21.1
14	Blue	Regulation of gene expression (chromatin activity)	1295	14.8
15	Pink	Lysosome, phagosome activity	621	14.4
16	Green	Cellular growth	918	13.9
17	Grey	No collective function (contains genes not associated with any module)	193	6.6
Total genes in network			11,950	

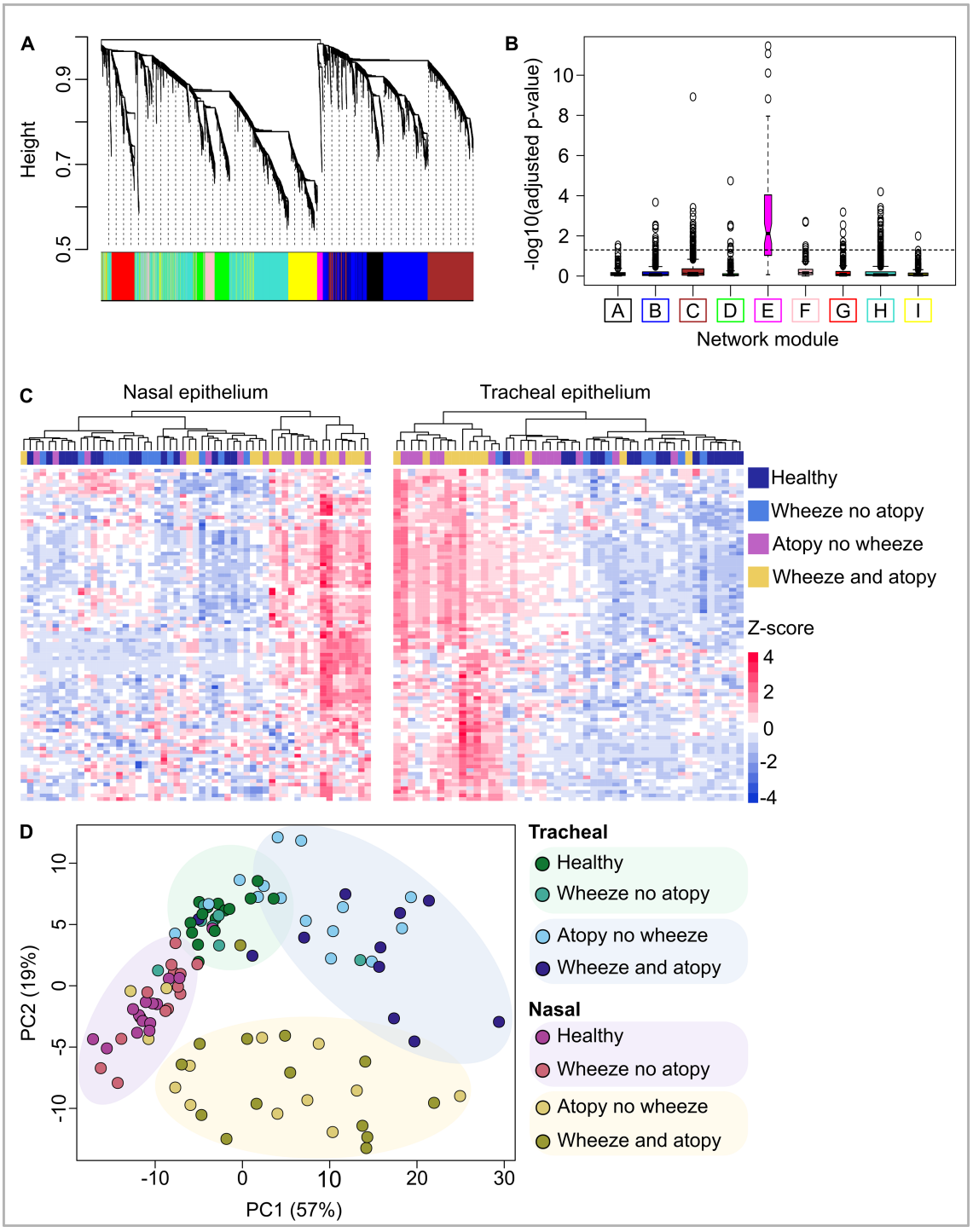
Table 4. Summary of differentially expressed genes between atopic and or asthmatics and non-atopic healthy controls in tracheal and nasal epithelium

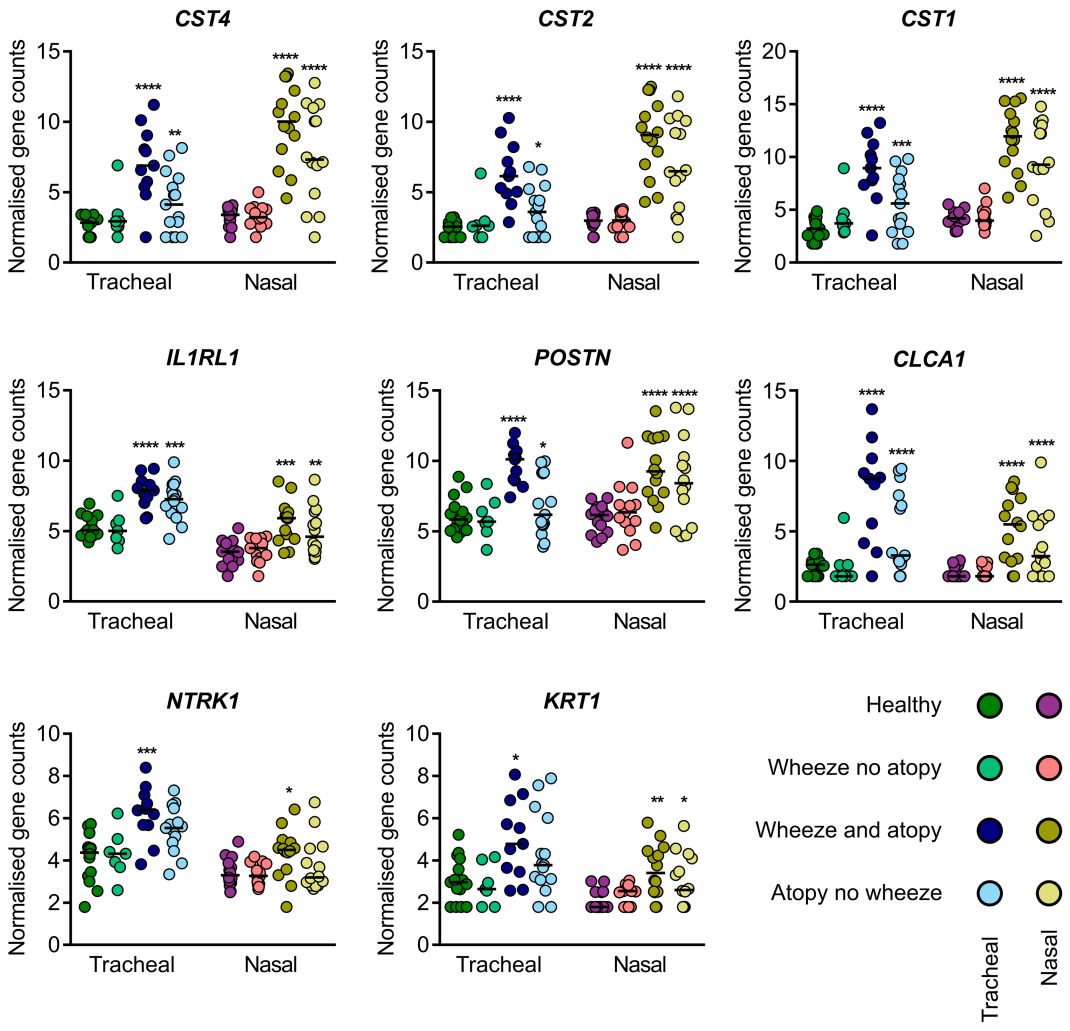
Contrast: description	**# of differentially expressed genes (adjusted p-value <0.05)**		
	Tracheal	**Nasal**	**Overlap**
Wheeze with atopy vs healthy controls: Effect of wheeze and asthma	**340** (98 down, 242 up)	**103** (13 down, 90 up)	**49** (2 down, 47 up)

Atopy only vs healthy controls: Effect of atopy only	**95** (55 down, 40 up)	**38** (1 down, 37 up)	**18** (1 down, 17 up)
Wheeze only vs healthy controls: Effect of wheeze only	**8** (3 down, 5 up)	**41** (25 down, 16 up)	**4** (1 down, 3 up)
Wheeze with atopy vs atopy only: Effect of wheeze in presence of atopy	**34** (23 down, 11 up)	**3** (1 down, 2 up)	**2** (1 down, 1 up)




A
Nasal network (8 modules)
345 338 92 2 58 45 91 83 8 5 681 555 292 597 49 267 230
16 34 28 0 10 22 33 21 0 0 11 15 6 35 8 93 39
3 9 14 0 8 3 24 55 2 1 4 3 0 4 12 36 7
10 251 70 0 4 21 31 463 8 0 6 2 1 13 15 18 34
42 496 97 9 93 275 423 217 335 76 61 22 201 326 69 552 551
3 8 22 2 225 5 37 0 0 0 8 1 0 0 0 35 4
13 69 161 3 101 146 731 12 13 9 11 7 0 9 25 141 33
11 33 36 120 83 104 400 8 13 3 7 5 8 11 15 153 20
50
40
30
20
10
0
-log10(p-value)
of overlap
Tracheal network (17 modules)
B
Consensus network (16 modules)
C
Median log2 fold change
(TEC vs NEC)
-1.0
-0.5
0.0
0.5
1.0
Cilium assembly / cell cycle
Cilium assembly / cell cycle
Protein folding
Gene expression
Protein binding
Immune response
Metabolism/ECM
Ribosome activity
Mitochondrial activity
Hormone signalling
Gene expresison
Regulation of transcription
Immune response (interferons)
Mitochondrial activity
Metabolism
Cellular growth
Higher expression
in nasal epithelium
Higher expression in
tracheal epithelium


A
Height
0.9
0.7
0.5
B
-log10(adjusted p-value)
10
8
6
4
2
0
A
B
C
D
E
F
G
H
I
Network module
C
Nasal epithelium
Tracheal epithelium
Healthy
Wheeze no atopy
Atopy no wheeze
Wheeze and atopy
Z-score
4
2
0
-2
-4
D
PC2 (19%)
10
5
0
-5
-10
-10
0
10
20
30
PC1 (57%)
Tracheal
Healthy
Wheeze no atopy
Atopy no wheeze
Wheeze and atopy
Nasal
Healthy
Wheeze no atopy
Atopy no wheeze
Wheeze and atopy


CST4
CST2
CST1
IL1RL1
POSTN
CLCA1
NTRK1
KRT1
Normalised gene counts
Tracheal
Nasal
Healthy
Wheeze no atopy
Wheeze and atopy
Atopy no wheeze
Tracheal
Nasal