Meta-analysis of airway epithelium gene expression in asthma

Yi-Hsuan Tsai[1], Joel S. Parker[1,2], Ivana V. Yang[3] and Samir N.P. Kelada [2,4]

Affiliations: [1]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA. [2]Dept of Genetics, University of North Carolina, Chapel Hill, NC, USA. [3]Dept of Medicine, University of Colorado, Aurora, CO, USA. [4]Marsico Lung Institute, University of North Carolina, Chapel Hill, NC, USA.

Correspondence: Samir N.P. Kelada, University of North Carolina, Department of Genetics, 120 Mason Farm Road, Chapel Hill, NC 27599, USA. E-mail: samir_kelada@med.unc.edu

ABSTRACT   Differential gene expression in the airway epithelium of patients with asthma *versus* controls has been reported in several studies. However, there is no consensus on which genes are reproducibly affected in asthma. We sought to identify a consensus list of differentially expressed genes (DEGs) using a meta-analysis approach.

We identified eight studies with data that met defined inclusion criteria. These studies comprised 355 cases and 193 controls and involved sampling either bronchial or nasal epithelium. We conducted study-level analyses, followed by a meta-analysis. Likewise, we applied a meta-analysis framework to the results of study-level pathway enrichment.

We identified 1273 DEGs, 431 of which had not been identified in previous studies. 450 DEGs exhibited large effect sizes and were robust to study population differences in age, sex, race/ethnicity, medication use, smoking status and exacerbations. The magnitude of differential expression of these 450 genes was highly similar in bronchial and nasal airway epithelia. Meta-analysis of pathway enrichment revealed a number of consistently dysregulated biological pathways, including putative transcriptional and post-transcriptional regulators.

In total, we identified a set of genes that is consistently dysregulated in asthma, that links to known and novel biological pathways, and that will inform asthma subtype identification.

Introduction

Multiple lines of evidence indicate that the airway epithelium plays an important role in the pathogenesis of asthma. Pattern recognition receptors, cytokines and alarmins secreted by airway epithelia play critical roles in the initiation of allergic inflammation [1]. Results from genome-wide association studies (GWAS) also implicate genes expressed by the airway epithelium, *e.g. IL33* and *TSLP* [2]. Protease-containing allergens (*e.g.* house dust mite) disrupt epithelial barrier integrity, and there is evidence of decreased barrier function in patients with asthma [3]. Lastly, the hallmark asthma phenotype of mucus hyper-secretion is caused by changes in the function and activity of secretory cells in the airway and submucosal glands [4].

Transcriptomic studies of airway epithelium have provided a snapshot of the state of this tissue after disease onset. WOODRUFF *et al.* [5] first reported on the identification of 22 genes that were differentially expressed between asthma cases and controls. This group later showed that the expression of just three genes (*POSTN*, *CLCA1* and *SERPINB1*) distinguished "T-helper cell 2 (Th2)-high" from "Th2-low" subtypes of asthma and was associated with response to inhaled corticosteroids [6]. Other studies have shown that changes in gene expression are mirrored by changes in the epithelial methylome [7, 8] and that distinct gene sets are associated with asthma exacerbations [9, 10]. Thus, the airway epithelium not only may be crucially involved in disease pathogenesis, but could also provide information about disease status (*i.e.* exacerbations) and therapeutic response.

Since the initial studies of gene expression in epithelial cells in asthma, more have followed, including studies with larger sample sizes and different demographics. While these more recent studies have contributed to our understanding of epithelial responses in asthma, with few exceptions there is no consensus on which genes and pathways are reproducibly affected in asthma. We therefore set out to utilise existing gene expression datasets to identify which genes were consistently dysregulated in the airway epithelium of asthma cases *versus* controls using a meta-analysis approach. In lieu of sampling bronchial epithelium, some studies used nasal epithelia, allowing us to assess the concordance of results between nasal and bronchial epithelia. We also used the asthma case–control gene expression data to evaluate enrichment of biological pathways in a similar meta-analysis framework.

Methods

Additional information on methods is available in the supplementary material.

Study identification

Case–control studies were identified in PubMed, National Centre for Biotechnology Information's Gene Expression Omnibus (GEO) and the European Molecular Biology Laboratory-European Bioinformatics Institute's ArrayExpress using search terms including "asthma", "airway epithelia", "bronchial epithelia" and "nasal epithelia" (or combinations thereof). In total, we identified 18 studies (supplementary table S1). Some publications contained overlapping study populations; thus, we chose one study from each set to avoid duplication (supplementary material for details).

Prior to reviewing the studies, we designated a set of study exclusion criteria. These were 1) lacking a case definition; 2) sample size less than five in any study group (cases or controls), which would limit power; or 3) unavailability of gene expression or covariate (age, sex and smoking) data. Covariate data were obtained either through the data repository or through contact with the study investigators. Application of the aforementioned exclusion criteria resulted in the exclusion of six studies (listed in supplementary table S1). After these exclusions (and removal of duplicates), our meta-analysis consisted of eight studies comprising 548 subjects (355 cases and 193 controls). All eight studies used brushes to sample airway epithelium: two studies sampled nasal epithelia while the rest sampled bronchi. All studies but one utilised microarray platforms. Descriptions of each study are provided in table 1.

Data processing, normalisation and annotation

The overall workflow for data acquisition, normalisation, study-level analysis and meta-analysis is shown in supplementary figure S1. Data were obtained either from GEO or directly from the study investigator. Affymetrix array data were normalised using robust multi-array averaging [11] and Agilent array data were quantile normalised and then log2 transformed (supplementary table S2). Normalised read counts from the single RNA-sequencing dataset (GSE85568) were also log2 transformed. In most cases, we used the array probe-gene annotation provided in GEO corresponding to the array platform used (supplementary table S2).

Study-level analyses

To identify outlier samples within each study, we used Bolstad's relative log expression method [12]. We then conducted an analysis to identify differentially expressed genes (DEGs) using LIMMA [13], adjusting

TABLE 1 Descriptive summary of studies included in gene expression meta-analysis

GEO ID	PubMed ID	Expression measurement platform	Epithelia source	Asthma[#, ¶]	Controls[#, +]	Total sample size[§]	Outliers[#]	Mean age (range)	Male %	Female %	Race/Ethnicity (Caucasian/ African American/ Hispanic/ Other)%	Covariates used in analysis	DEGs[#] (up/down)
GSE19187	22005912	Affymetrix Human Gene 1.0 ST Array	Nasal	13	11	24	0	11 (6–16)	53	47	100/0/0/0	Age, sex	116 (104/12)
GSE67472	25611785	Affymetrix Human Genome U133 Plus 2.0 Array	Bronchial	62	43	105	1	35 (20–68)	49	51	51/15/19/14	Age, sex, race, array batch	6504 (4112/2392)
GSE63142	25338189	Agilent-014850 Whole Human Genome Microarray 4×44K	Bronchial	125	27	152	6	36 (14–63)	36	64	62/29/0/9	Age, sex, race	1078 (662/416)
GSE65205	27745942	Agilent-028004 SurePrint G3 Human GE 8×60K Microarray	Nasal	36	33	69	3	11 (9–12)	49	51	0/100/0/0	Age, sex, study site	147 (59/88)
GSE104472	28294656	Agilent-028004 SurePrint G3 Human GE 8×60K Microarray	Bronchial	12	12	24	1	36 (24–74)	42	58	100/0/0/0	Age, sex	20 (13/7)
GSE64913	28045928	Affymetrix HG U133 plus 2.0	Bronchial	13	23	36	0	31 (19–63)	59	41	100/0/0/0	Age, sex, array batch	9 (7/2)
GSE89809	28933920	Affymetrix HT HG-U133+	Bronchial	38	16	54	3	38 (20–67)	57	43	100/0/0/0	Age, sex, smoking	2 (2/0)
GSE85568	27942592	RNA-seq	Bronchial	56	28	84	3	39 (19–62)	32	68	34/60/0/5	Age, sex, race, smoking	17 (9/8)

DEG: differentially expressed gene. [#]: at false discover rate q<0.1; [¶]: n=355; [+]: n=193; [§]: n=548.

at a minimum for age, sex and race, but in some cases also smoking status and array hybridisation batch (table 1). For each study, p-values were adjusted using the false discovery rate (FDR) correction method [14], and we recorded the t statistic, beta, p-value, log fold change, Cohen's D and Hedges' adjusted g for use in the meta-analysis.

Meta-analysis

We estimated summary effect sizes for each gene using an inverse variance model as advocated by RAMASAMY *et al.* [15]. Specifically, we calculated pooled effect sizes using Hedges' adjusted g and included the use of a random effects model to account for between-study heterogeneity. This analysis was carried out using the R package meta [16]. We declared genes as differentially expressed (DE) if the meta-analysis FDR-adjusted p-value (q-value) was <0.05 and the gene was evaluated in at least two of the eight studies.

Overlap of asthma DEGs with GWAS hits

To assess whether any of the DEGs we identified by meta-analysis were also implicated by GWAS, we merged our meta-analysis results with data from the National Human Genome Research Institute-European Bioinformatics Institute Catalog of published GWAS [17, 18]. To be inclusive, we applied a GWAS p-value threshold of 1×10^{-7}. We used the gene reported by the original study as the key identifier, acknowledging that the gene implicated by GWAS is not necessarily the causal gene. If the same gene was reported from multiple GWAS, we report the single most significant result.

Pathway enrichment meta-analysis

We evaluated pathway enrichment in each study using Gene Set Variation Analysis (GSVA) [19]. Briefly, GSVA works by estimating a gene set score in an unsupervised way by calculating an enrichment statistic for each gene set in each study (supplementary material). In this way, the gene by sample matrix is transformed to a matrix of gene set by sample. The gene set by sample matrix was then used for study-level differential expression analysis, followed by meta-analysis using the same methods as described above for individual genes. Gene sets for analysis were collected from Enrichr (http://amp.pharm.mssm. edu/Enrichr/) [20, 21] and include Kyoto Encyclopedia of Genes and Genomes (KEGG) and Biocarta pathways, transcription factor targets from the Encyclopedia of DNA Elements (ENCODE) and many others (supplementary material).

Results

Study-level analyses

We obtained gene expression and covariate data for 355 cases and 193 controls from eight studies (table 1). Demographic variables differed considerably between studies. For example, two studies included only children, whereas the remaining studies largely included adults of a wide age range (20–74 years). Across studies, the percentage of study subjects that were female ranged from ~40% to 70%. While most study subjects were Caucasian, some studies involved African-Americans, Hispanics, Asians or other racial/ethnic categories. Except for two studies, most studies involved exclusively non-smoking subjects.
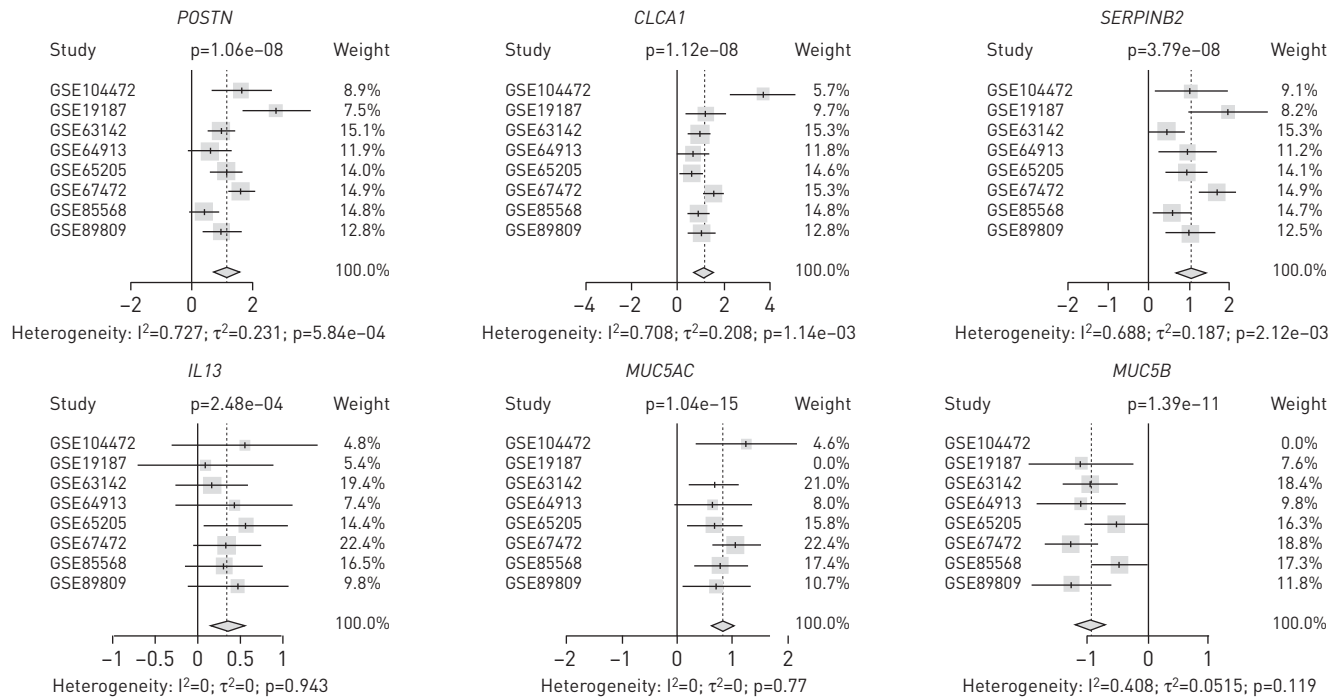
For each study, we conducted an analysis to identify DEGs after the exclusion of outliers (supplementary figure S2). Because different studies utilised different array platforms (or RNA-sequencing) with different numbers of genes tested, we compared the eight studies by plotting the fraction of genes identified as DE (supplementary figure S3, supplementary table S2). These values ranged from 0.01% to 27%, with no obvious correlation between sample size and the fraction of DEGs detected.
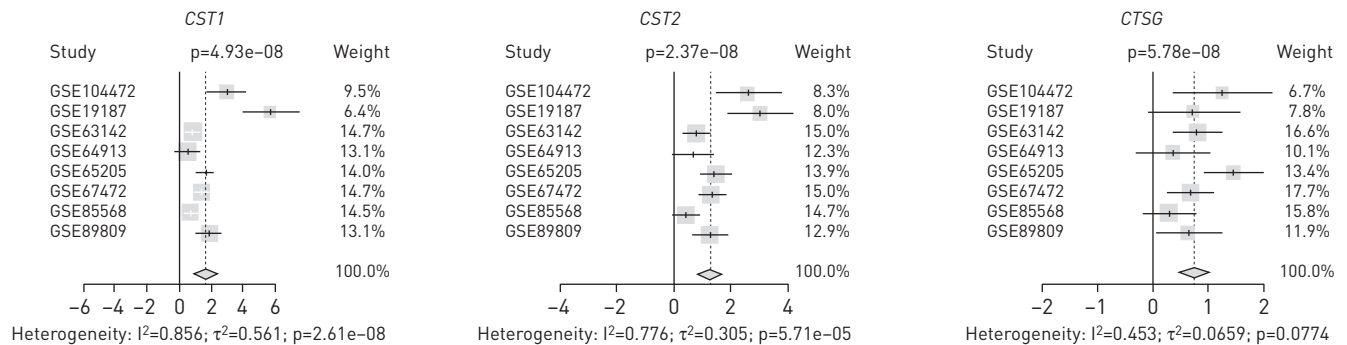
Meta-analysis

We then applied a meta-analysis approach, using a random effects model to allow for heterogeneity in effect sizes across studies. We identified 1273 genes that were DE with q-values <0.05 (figure 1, supplementary table S3), the majority (64%) of which were upregulated. Notably, more than one third of these DEGs (n=431, 34%) were not identified as DE in any single study (at FDR q<0.05), highlighting the utility of the approach to reveal genes of interest. In terms of effect size, *CST1* (upregulated) and *APELA* (downregulated) were the most extreme genes; in terms of statistical significance, *CEACAM5* (upregulated) and *C3* (downregulated) were the most affected genes.

As expected, the three Th2 biomarker genes (*POSTN*, *CLCA1* and *SERPINB1*) were consistently upregulated, though with some heterogeneity in effect size across studies (figure 2). Likewise, genes positively associated with mucus production (*e.g. IL13*, *FOXA3* and *MUC5AC*) were significantly upregulated, while *MUC5B* and *FOXA2* were downregulated. Genes related to endoplasmic reticulum stress (*e.g. AGR2* and *ERN1*), which has been linked to mucin hyper-secretion [22, 23], were also upregulated, though *XBP1* was not (supplementary figure S4).

FIGURE 1 Differentially expressed genes identified by meta-analysis. Genes are colour-coded by the number of studies (k) in which they were evaluated, and those with adjusted p-values $<1\times10^{-5}$ and pooled effect size >0.7 are labelled.

We noted an obvious enrichment of genes involved in protease/antiprotease pathways, including serine protease inhibitors (*SERPINB2*, *SERPINB4*, *SERPINB8* and *SERPINB10*), cathepsins (*CTSC* and *CTSG* were upregulated whereas *CTSL* was downregulated) and cystatins (*CST1*, *CST2*, *CST3*, *CST4*, *CST6* and *CSTA*) (figure 2, supplementary figure S4). In contrast, genes involved in barrier function (including *CLDN1*, *CLDN18*, *CLDN4*, *TJP1*, *TJP2*, *TJP3*, *OCLN* and *CDH1*) were not DE, with the sole exception of *TJP1* (*ZO1*). Six genes reflective of recruited leukocytes were DE, including *CCL26* and *CLC* (markers of eosinophils) and *TPSAB1*, *CPA3*, *MS4A2* and *HDC* (markers of mast cells and/or basophils) (supplementary figure S4).

Seventeen genes implicated by GWAS were in the list of DEGs (supplementary table S4). While overall there was no over-representation of upregulated *versus* downregulated genes in this list, the five genes showing the most differential expression in terms of effect size (*LRRC8A*, *ALOX15*, *IL18R1*, *IL1RL1*, *ADAMTS9*) were all upregulated. Conversely, all three HLA genes (*HLA-DOA*, *HLA-DPA1*, *HLA-DRA*) were downregulated. Surprisingly, we found decreased expression of the asthma-associated gene *CHI3L1* [24] among cases (figure 2). Genes identified by positional cloning, including *ADAM33*, *DPP10*, *PCDH1*, *HLA-G* and *SPINK5,* were not DE, nor was *CDHR3* [25], a gene associated with asthma exacerbations.

A large number of genes (n=450) were consistently DE across many of the eight studies and had large effect sizes, which we defined as pooled effect size (Hedges' g) values >0.5 (figure 3, supplementary table S5). These findings indicate that differential expression of these genes is robust to differences in age, sex, smoking status, race/ethnicity, exacerbations and tissue type (bronchial *versus* nasal). We further examined the commonality of differential gene expression between bronchial and nasal epithelia by comparing the pooled effect size from the six studies involving bronchial epithelia *versus* the pooled effect size from the two studies involving nasal epithelia (figure 4). This analysis revealed that the pattern of differential gene expression across tissue types was completely consistent for these 450 genes, and the magnitude of differential expression for most genes was only marginally decreased in nasal compared to bronchial epithelia (slope=0.91, p<0.01 for test that slope=1), though *CST1* was more pronounced in nasal epithelia than in bronchial epithelia.

Heterogeneity

While the consistency of differential gene expression signal in asthma cases *versus* controls was clear for many genes, in particular the Th2 marker genes *POSTN*, *CLCA1* and *SERPINB1*, this level of analysis obscures potential heterogeneity that could inform disease mechanism and treatment [6]. To examine this, we plotted the distribution of aggregate expression scores for *POSTN*, *CLCA1* and *SERPINB1* (figure 5). We found that while on average cases had higher levels of expression of these genes in aggregate, within each study the distribution ranged widely and was in some cases bimodal (*e.g.* GSE67472). Additionally, in some studies there were cases with lower Th2 gene expression than controls (*e.g.* GSE63142).
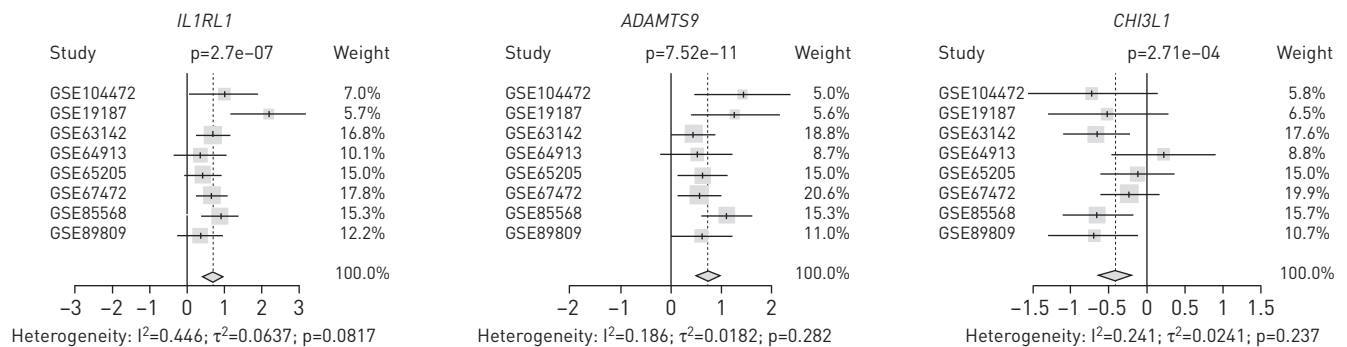
FIGURE 2 Differential gene expression across studies for emblematic genes. Forest plots for a) genes related to Th2-mediated inflammation (*POSTN*, *CLCA1* and *SERPINB1*) and mucin production (*IL13*, *MUC5AC* and *MUC5B*); b) protease antiprotease balance (*CST1*, *CST2* and *CTSG*); and c) genes with variants associated with asthma (*IL1RL1*, *ADAMTS9* and *CHI3L1*). White whiskers (for *CST1*) indicate that the confidence interval is completely contained within the box representing the point estimate. Additional forest plots are provided in supplementary figure S4.

Effect of steroids

The lack of data on medication use precluded a systematic analysis of inhaled corticosteroid (ICS) effect. Hence, we compared our list of DEGs to previous studies which examined the effect ICSs have on airway
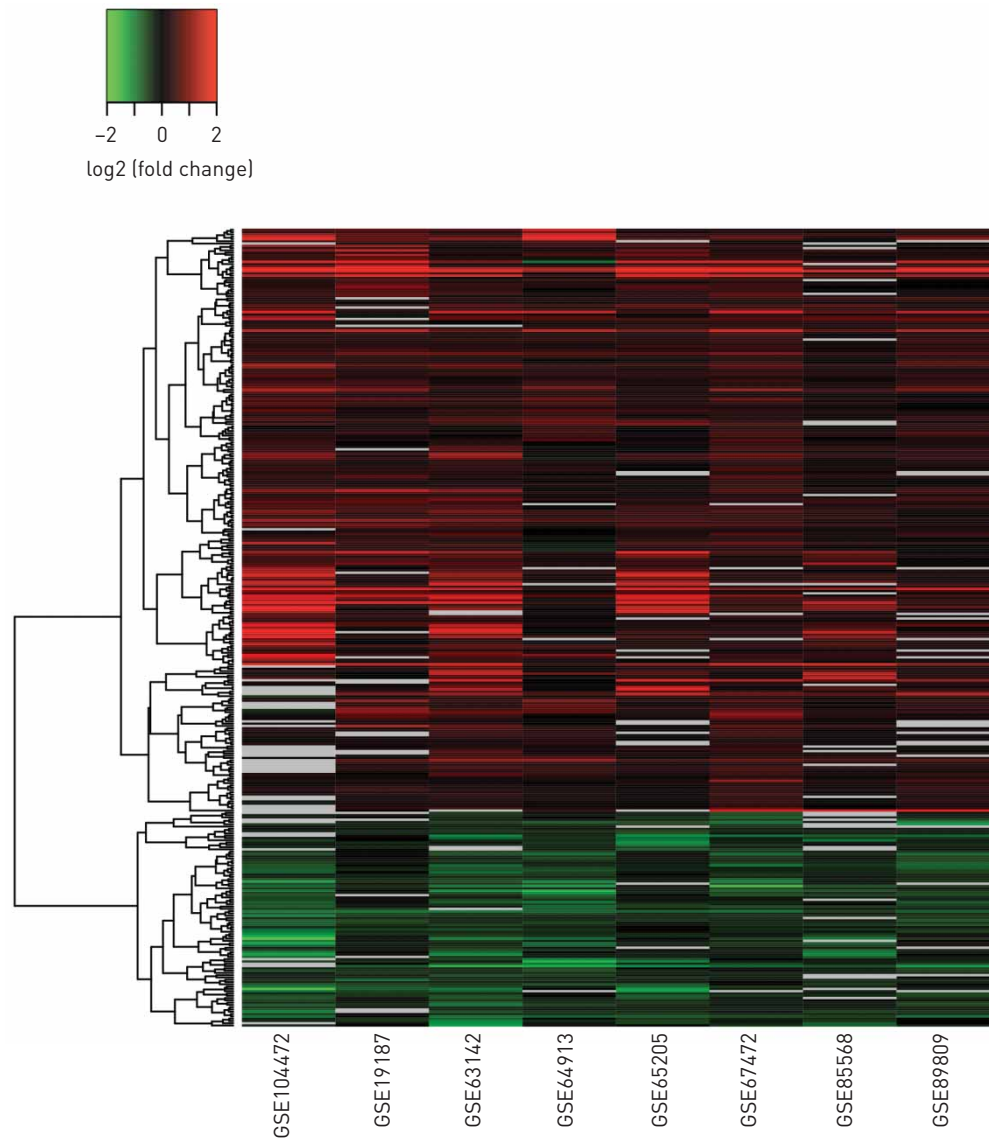
FIGURE 3 Consistency of differential expression for differentially expressed genes (DEGs) with large effect size. DEGs with adjusted p-values <0.05 and pooled effect size >0.5 are shown. Log2 fold change values >2 or <−2 were set to 2 or −2. Missing values are shown in grey. 450 genes were identified with some degree of missingness across studies. For ease of viewing, the 349 genes with the least missingness are shown. A similar heat map for all 1273 DEGs is provided in supplementary figure S5.

epithelia transcriptomes [5, 26], which in aggregate identified 111 genes. This list of ICS-responsive genes included 15 asthma DEGs, including *ALOX15B*, *POSTN*, *CLCA1*, *SERPINB1* and *CST1* (supplementary table S6).

Pathway meta-analyses

To examine the effect of differential gene expression at the level of biological pathways, which could reveal additional disease-associated effects that are not apparent at the single-gene level, we conducted a meta-analysis of pathway enrichment. First, we conducted study-level pathway analysis using the GSVA method [19] to examine variation in the expression of defined gene sets (*e.g.* KEGG pathways and other predefined gene sets) between cases and controls in each study. In this analysis, genes that were not called DE in either the single study analysis or meta-analysis could still contribute to pathway enrichment. That is, because GSVA works by combining expression scores across a predefined set of genes, modest differences in the expression of genes that would not reach statistical significance (for each gene individually) can still cumulatively result in differential expression of a gene set. Following single study analysis, we conducted a meta-analysis of pathway expression scores across all eight studies to identify
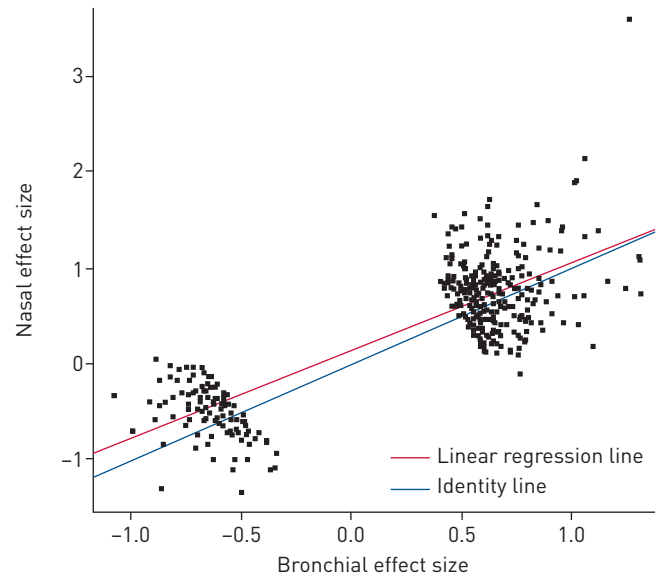
FIGURE 4 Similarity of differential expression in bronchial *versus* nasal epithelium. A comparison of pooled effect sizes for the top 450 differentially expressed genes is shown, along with best-fit linear regression line (red) and the line of identity (blue; *i.e.* if nasal=bronchial). The outlier gene at the top right is *CST1*.

consistently altered pathways (figure 6, supplementary table S7, supplementary table S8). Some of the most prominent pathways with increased expression in cases were linked to mucin synthesis and post-translational modification of mucins. Pathways related to the synthesis of prostaglandins, thromboxanes and eicosatetraenoic acid derivatives also showed increased expression. In controls, interferon gamma signalling, tryptophan metabolism, NOTCH and Hedgehog signalling were among the more highly expressed pathways. Additionally, this analysis identified TCF12, XBP1 and FOSL2 as putative transcriptional regulators of upregulated genes, and miR-380–5p as a post-transcriptional regulator of genes that are more highly expressed in controls. We also queried three protein interaction databases and identified 335 putative protein interaction partners (at FDR q-values <0.05) of the asthma DEGs (supplementary table S9), the majority of which have not previously been linked to asthma.

Discussion

Our meta-analyses of eight studies revealed that 1273 genes, roughly 5% of the protein-coding genes in the human genome, are consistently dysregulated in the airway epithelium of patients with asthma, and more than one third of these DEGs had large effect sizes. The robustness of our meta-analysis results is largely a function of the similarity of differential gene expression across each of the eight studies; the fact that the eight studies utilised different expression measurement platforms makes the results particularly convincing. Part of the utility of this analysis is the identification of genes that did not stand out in any single study, which comprised 34% of the total number of DEGs. *PPP1R3B* is one such example. This gene was nominally differentially expressed in four out of the eight studies but after correction for multiple testing was not called DE in any single study. This gene is associated with lipid metabolism [27] and Alzheimer's disease [28], but has not previously been linked to asthma.

The results of our pathway meta-analysis clearly indicate that multiple pathways are consistently dysregulated in asthma, though we cannot infer whether these pathways are causally related to the development of disease or are a consequence of disease. Genes related to Th2-mediated inflammation, MUC5AC production and post-translational modification of mucins were consistently and strongly upregulated in asthma cases. The pathway analysis also implicated endoplasmic reticulum stress (IRE1 and XBP1 pathways), which is a known consequence of exaggerated mucin production [22, 23]. While XBP1 was identified as a putative regulator of asthma DEGs, *XBP1* gene expression did not differ by case–control status, suggesting that activation of XBP1 function may occur independently of *XBP1* transcription. In contrast to heightened mucin gene expression, genes involved in NOTCH signalling were more highly expressed in controls. Collectively, these particular pathway enrichment results seem to point to both alterations in mucus production and changes in goblet *versus* ciliated cell numbers in the airways.

The clear enrichment of genes from the serpin, cystatin and cathepsin gene families in the list of DEGs highlights the role of protease–antiprotease imbalance and the consequences thereof, perhaps most
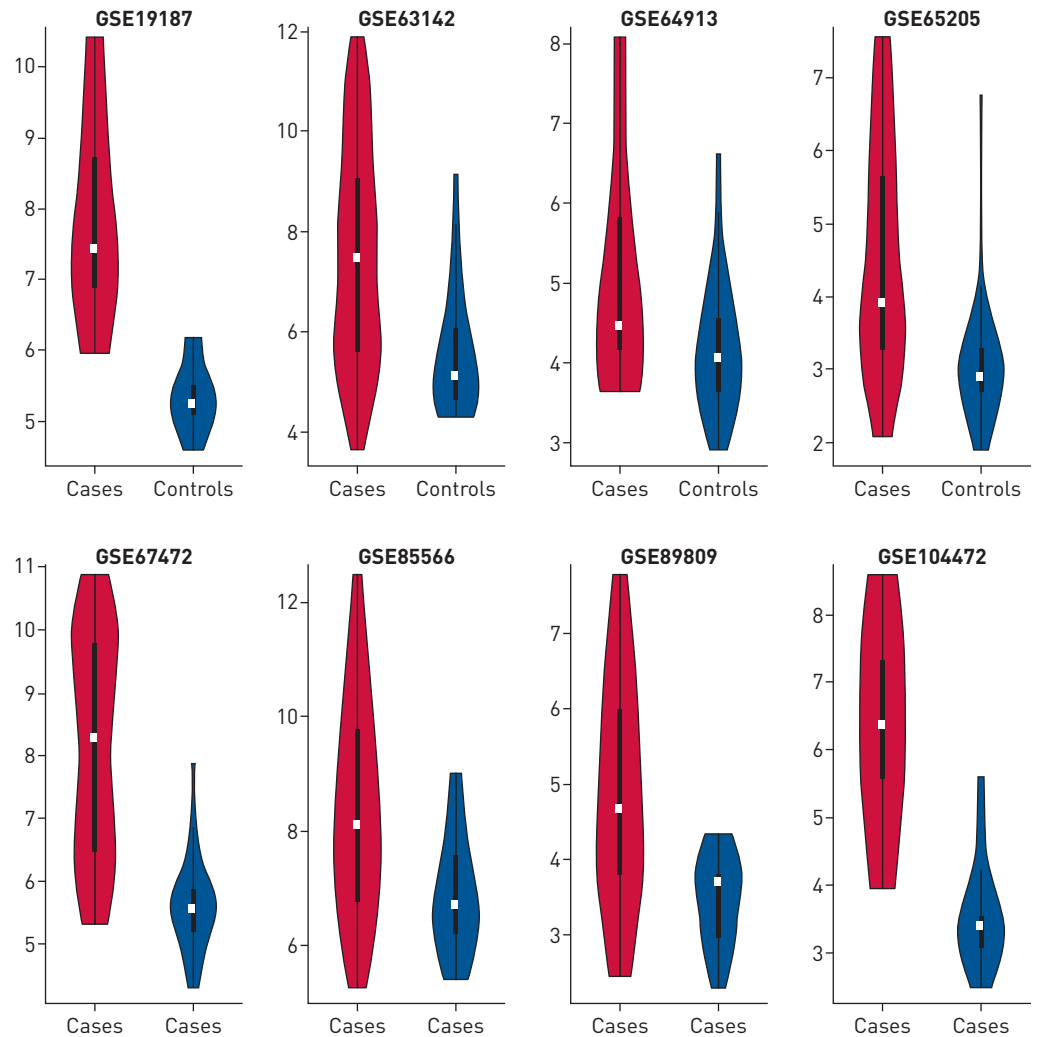
FIGURE 5 Variation in Th2 marker gene expression among studies and samples. The sum of gene expression for *POSTN*, *CLCA1* and *SERPINB2* (in log2 units) is plotted as a function of case–control status across each study. Cases are shown in red and controls in blue. In these violin plots, white dots denote median, the thick black bars correspond to the first to third quartiles of the distribution, and the bottom and top whiskers correspond to the first quartile–1.5×IQR and the third quartile+1.5×IQR, respectively, where IQR=interquartile range.

prominently extracellular matrix remodelling. Previous findings suggest that some of the protease expression is likely derived from recruited white blood cells, such as mast cell tryptase and carboxypeptidase, and neutrophil-derived cathepsin G and elastase. It is worth noting that heightened expression of such proteases has also been linked to altered biophysical properties of mucins in the airway, potentially leading to mucus plugging [29]. The upregulation of cystatins, which are protease inhibitors, may also be indicative of a compensatory response to exposure to protease-containing allergens such as the house dust mite, as has been suggested in the case of eosinophilic chronic rhinosinusitis [30]. However, we did not find strong evidence in support of a barrier defect in the asthmatic airway epithelium, as one would expect based on the known effect of proteases on epithelial integrity [3]. Nor did we find evidence of a pervasive effect of steroids on DEGs, but this result was based on a limited analysis of previous data and therefore a more robust analysis is merited.

One of our key findings was that dysregulated gene expression in asthma is reflected in the nasal epithelium at a magnitude that is highly similar to the bronchial epithelium. This result corroborates earlier findings of similarities in gene expression [31] and DNA methylation [8] based on comparisons of matched samples. Because our results are based on comparisons across studies, we conclude that this commonality of expression patterns is generalisable to multiple study populations. Notably, this commonality of differential expression includes genes that likely represent leukocyte infiltration, *i.e.*
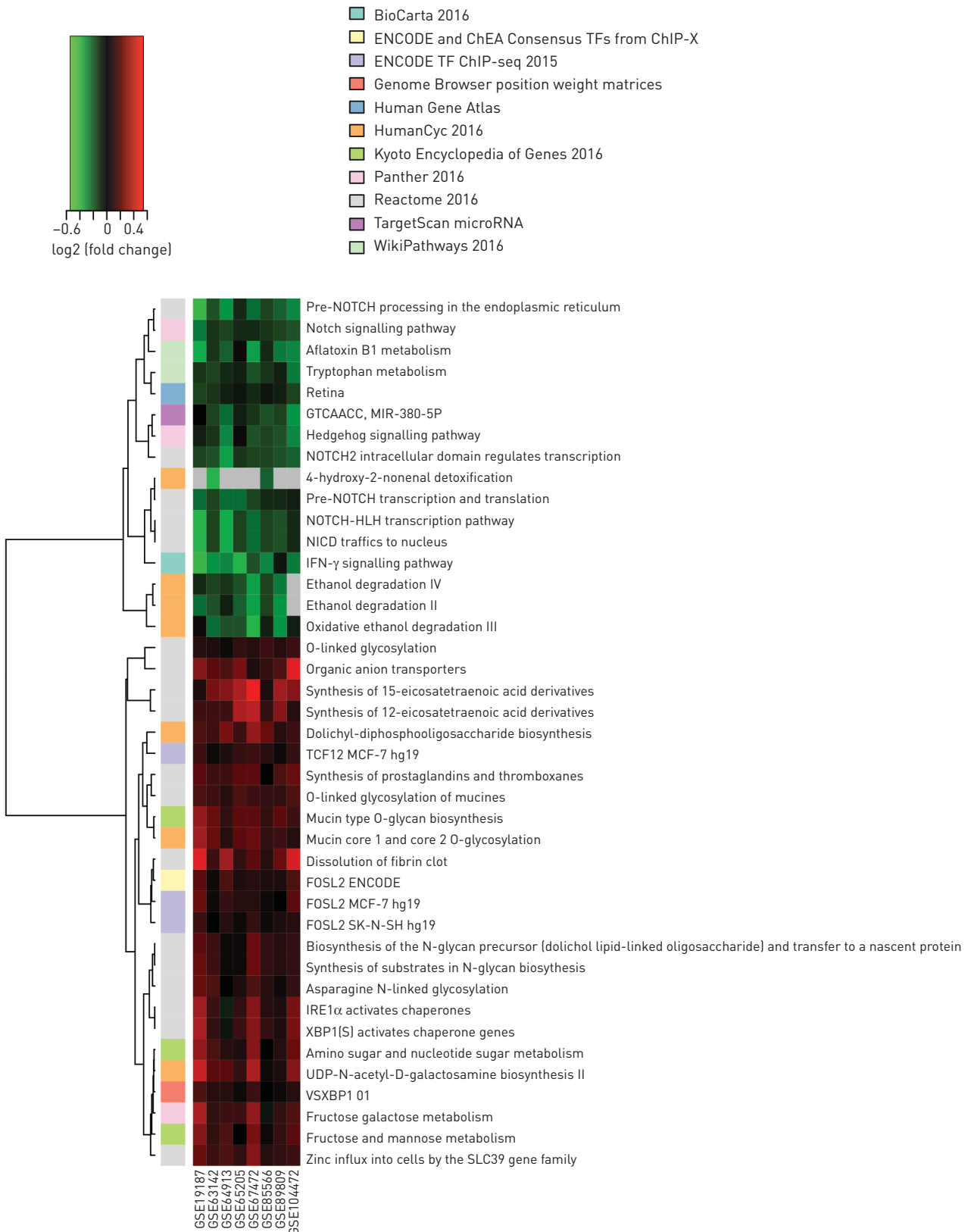
FIGURE 6 Consistency of differential pathway expression revealed by meta-analysis. A heat map of fold change (log2 scale) for top pathways identified by meta-analysis of pathway enrichment across eight studies is shown. In some cases, pathway names have been truncated for ease of viewing. The full list of pathway names is provided in supplementary table S7. In supplementary table S8, we provide a cross referencing of differentially expressed genes to pathways. ChEA: ChIP Enrichment Analysis; TF: transcription factor; ChIP: chromatin immunoprecipitation.

CCL26, *CLC*, *TPSAB1*, *CPA3*, *MS4A2* and *HDC*). Heightened expression of *CLC* and *HDC* has also been observed in allergic rhinitis [32] and eosinophilic esophagitis [33] (more details are provided in the supplementary material), suggesting this is a common feature of eosinophilic inflammation across tissues.

We found 17 genes implicated by GWAS that were DE in the airway epithelium of patients with asthma. Caution must be exerted when interpreting this result because the assignment of a gene from an association signal (in the GWAS) based on physical proximity is not proof positive of causality. With regards to *IL1RL1* and *CHI3L1*, however, there is ample evidence from other studies that these two genes are causally related to asthma. For these two genes, our meta-analysis indicated an apparent discrepancy between the directionality of gene expression and case status and how genetic variants at these loci affect gene expression in relation to disease risk. For *IL1RL1*, expression is upregulated in asthma, yet the risk variants associated with asthma (as well as serum IgE and peripheral blood eosinophil counts) [34] are associated with lower gene expression [35]. This suggests that asthma status may cause an increase in *IL1RL1* gene expression in airway epithelia and/or that recruited white blood cells present in the asthmatic airway epithelium express *IL1RL1*. For *CHI3L1*, which encodes YKL-40, genetic variants associated with increased risk of asthma are strongly associated with higher levels of serum YKL-40 [24] and gene expression in whole lung tissue [36], and YKL-40 is expressed in the bronchial epithelium, macrophages and neutrophils of patients with severe asthma [37]. Yet, our results show that the gene was expressed at lower levels in cases *versus* controls overall. It is unclear whether this discrepancy is attributable to variation in cell type composition of the airways or disease severity, or whether there is a distinction between gene expression in the airway epithelium before and after disease onset that may affect this relationship. Lastly, genetic variants downstream of *ALOX15* are associated with asthma [38] and this gene was also significantly and highly upregulated in patients with asthma. This gene encodes arachidonate 15-lipoxygenase, an enzyme that converts arachidonic acid into several eicosatetraenoic acid species (which was also identified in the pathway meta-analysis), some of which have been shown to affect mucin production by bronchial epithelial cells [39, 40]. Thus, the expression and GWAS data contribute to a biologically coherent argument in favour of a role for this gene in asthma.

While our results clearly indicate that a sizeable number of genes are DE in asthma, this is not to say there are not differences between studies or heterogeneity among patients with asthma. Indeed, some studies identified a comparatively large number of DEGs compared to others, and this was not simply a function of sample size. This result suggests that real biological differences between study populations led to the observed differences in gene expression signals. It seems possible that the presence of different asthma subtypes could lead to either increased or decreased power. The presence of Th2-low asthma cases, for example, could act to dilute differences between cases and controls for Th2-related genes. Alternatively, Th2-low cases could provide power to detect non-Th2-related expression differences. Variation in the three-gene marker of Th2-mediated inflammation (*POSTN*, *CLCA1* and *SERPINB1*) among cases was indeed present. In at least three studies, the distribution of aggregate Th2-gene expression was bimodal, and some patients with asthma had lower expression of these genes than any control. It is clearly imperative that the genetic and clinical correlates of this asthma subtype be identified. More broadly, the use of gene expression data to identify biological underpinnings of asthma subtypes (or endotypes) is a promising approach [6, 41]. This approach will depend in large part on the availability of study subject data (clinical and/or biomarker) that can be used to assess correlations with gene expression. The identification of 1273 consistently dysregulated genes in the airways of patients with asthma represents one starting point for these analyses in the future.

References

1   Hammad H, Lambrecht BN. Barrier epithelial cells and the control of type 2 immunity. *Immunity* 2015; 43: 29–40.
2   Ortiz RA, Barnes KC. Genetics of allergic diseases. *Immunol Allergy Clin North Am* 2015; 35: 19–44.
3   Georas SN, Rezaee F. Epithelial barrier function: at the front line of asthma immunology and allergic airway inflammation. *J Allergy Clin Immunol* 2014; 134: 509–520.
4   Fahy JV, Dickey BF. Airway mucus function and dysfunction. *N Engl J Med* 2010; 363: 2233–2247.
5   Woodruff PG, Boushey HA, Dolganov GM, *et al.* Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc Natl Acad Sci USA* 2007; 104: 15858–15863.

6   Woodruff PG, Modrek B, Choy DF, *et al.* T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med* 2009; 180: 388–395.

7   Yang IV, Pedersen BS, Liu AH, *et al.* The nasal methylome and childhood atopic asthma. *J Allergy Clin Immunol* 2017; 139: 1478–1488.

8   Yang IV, Richards AJ, Davidson E, *et al.* The nasal methylome: a key to understanding allergic asthma. *Am J Respir Crit Care Med* 2017; 195: 829–831.

9   Guajardo JR, Schleifer KW, Daines MO, *et al.* Altered gene expression profiles in nasal respiratory epithelium reflect stable versus acute childhood asthma. *J Allergy Clin Immunol* 2005; 115: 243–251.

10   Giovannini-Chami L. Distinct epithelial gene expression phenotypes in childhood respiratory allergy. *Eur Respir J* 2012; 39: 1197–1205.

11   Irizarry RA. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; 4: 249–264.

12   Bolstad BM, Collin F, Brettschneider J, *et al.* Quality assessment of Affymetrix GeneChip data. *Bioinforma Comput Biol Solut Using R Bioconductor* 2005; 33–47.

13   Ritchie ME, Phipson B, Wu D, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.

14   Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995; 57: 289–300.

15   Ramasamy A, Mondry A, Holmes CC, *et al.* Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med* 2008; 5: 1320–1332.

16   Schwarzer G, Carpenter JR, Rücker G. Meta-Analysis with R. First. Heidelberg, Springer, 2015.

17   GWAS Catalog: The NHGRI-EBI Catalog of published genome-wide association studies. www.ebi.ac.uk/gwas/ Date last accessed: January 31, 2018. Date last updated: April 4, 2018.

18   Welter D, MacArthur J, Morales J, *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014; 42: D1001–D1006.

19   Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013; 14: 7.

20   Chen EY, Tan CM, Kou Y, *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013; 128: 14.

21   Kuleshov MV, Jones MR, Rouillard AD, *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016; 44: W90–W97.

22   Martino MB, Jones L, Brighton B, *et al.* The ER stress transducer IRE1β is required for airway epithelial mucin production. *Mucosal Immunol* 2013; 6: 639–654.

23   Schroeder BW, Verhaeghe C, Park SW, *et al.* AGR2 is induced in asthma and promotes allergen-induced mucin overproduction. *Am J Respir Cell Mol Biol* 2012; 47: 178–185.

24   Ober C, Tan Z, Sun Y, *et al.* Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. *N Engl J Med* 2008; 358: 1682–1691.

25   Bønnelykke K, Sleiman P, Nielsen K, *et al.* A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* 2014; 46: 51–55.

26   Leigh R, Mostafa MM, King EM, *et al.* An inhaled dose of budesonide induces genes involved in transcription and signaling in the human airways: enhancement of anti- and proinflammatory effector genes. *Pharmacol Res Perspect* 2016; 4: e00243.

27   Wu Y, Waite LL, Jackson AU, *et al.* Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet* 2013; 9: e1003379.

28   Kamboh MI, Demirci FY, Wang X, *et al.* Genome-wide association study of Alzheimer's disease. *Transl Psychiatry* 2012; 2: e117.

29   Innes AL, Carrington SD, Thornton DJ, *et al. Ex vivo* sputum analysis reveals impairment of protease-dependent mucus degradation by plasma proteins in acute asthma. *Am J Respir Crit Care Med* 2009; 180: 203–210.

30   Kouzaki H, Matsumoto K, Kikuoka H, *et al.* Endogenous protease inhibitors in airway epithelial cells contribute to eosinophilic chronic rhinosinusitis. *Am J Respir Crit Care Med* 2017; 195: 737–747.

31   Poole A, Urbanek C, Eng C, *et al.* Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. *J Allergy Clin Immunol* 2014; 133: 670–678.e12.

32   Imoto Y, Tokunaga T, Matsumoto Y, *et al.* Cystatin SN upregulation in patients with seasonal allergic rhinitis. *PLoS One* 2013; 8: e67057.

33   Blanchard C. Eotaxin-3 and a uniquely conserved gene- expression profile in eosinophilic esophagitis. *J Clin Invest* 2006; 116: 536–547.

34   Savenije OEM, Kerkhof M, Reijmerink NE, *et al.* Interleukin-1 receptor-like 1 polymorphisms are associated with serum IL1RL1-a, eosinophils, and asthma in childhood. *J Allergy Clin Immunol* 2011; 127: 750–756.

35   Gordon ED, Palandra J, Wesolowska-Andersen A, *et al.* IL1RL1 asthma risk variants regulate airway type 2 inflammation. *JCI Insight* 2016; 1: 1–13.

36   The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; 348: 648–660.

37   Chupp GL, Lee CG, Jarjour N, *et al.* A chitinase-like protein in the lung and circulation of patients with severe asthma. *N Engl J Med* 2007; 357: 2016–2027.

38   Ferreira MA, Vonk JM, Baurecht H, *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet* 2017; 49: 1752–1757.

39   Jayawickreme SP, Gray T, Nettesheim P, *et al.* Regulation of 15-lipoxygenase expression and mucus secretion by IL-4 in human bronchial epithelial cells. *Am J Physiol* 1999; 276: L596–L603.

40   Zhao J, Maskrey B, Balzar S, *et al.* Interleukin-13-induced MUC5AC is regulated by 15-lipoxygenase 1 pathway in human bronchial epithelial cells. *Am J Respir Crit Care Med* 2009; 179: 782–790.

41   Modena BD, Tedrow JR, Milosevic J, *et al.* Gene expression in relation to exhaled nitric oxide identifies novel asthma phenotypes with unique biomolecular pathways. *Am J Respir Crit Care Med* 2014; 190: 1363–1372.