Supplemental Material for Parker et al.

Supplementary Methods

Comments on data aggregation and covariates

After the initial identification of studies, we noted that there was overlap in study populations examined in some studies, thus we chose one study from each set. For one group of such studies [1–3], we chose to use the largest study from this group because it contained most or all of the data from prior studies. For the second set of such studies[4–6], based on subjects in the Severe Asthma Research Program (SARP), we used the study for which airway epithelial data generation was best described[6].

Also, we did not include in vitro studies of cultured epithelia from cases vs. controls because preliminary analyses indicated substantially different gene expression patterns between in vivo and in vitro samples (not shown).

Case-control status

For the study by Giovanini-Chami et al. [7], we combined both types of asthma cases (controlled and uncontrolled) together. Likewise, for the study Modena et al. (GSE63142) [6], we combined data from severe asthmatics and mild-to-moderate asthmatics in the analysis of differential gene expression.

Smoking
For the majority of studies, subjects were non-smokers. However, for two studies, GSE85568 [8] and GSE89809 (unpublished), some study subjects smoked and hence smoking status was included in differential gene expression analyses for these studies.

Race
Not all studies reported data on the race/ethnicity of their study subjects. Rather than exclude these studies, we made the assumption that cases and controls from each study were reasonably well matched with respect to race/ethnicity. For GSE89809, we excluded one subject self identified as African because this was the only African subject in the study and analysis including this subject would not have allowed us to make meaningful inferences about the effect of race/ethnicity. For GSE63142, three subjects lacked race data and hence were excluded from the dataset.

Medication use
Most studies did not provide data on medication use in GEO. Therefore the effect of medications was not examined in the meta-analysis. Instead, we compared our list of DEGs to previous studies which examined the effect ICS on airway epithelia transcriptomes [9, 10]. From these two studies, we manually compiled a list of genes that were reported to be differentially expressed in either study, totaling 111 genes. We then examined overlap between these 111 genes and the genes from our meta-analysis.

Array probe annotation
In most cases, we used the array probe-gene annotation provided in GEO corresponding to the
array platform used. However, for the unpublished study by Yang et al. (NJH), we used Biomart to convert probes to genes using the “Agilent_sureprint_g3_ge_8x60k_v2” annotation. If multiple probes mapped to one gene, the probe with highest standard deviation was used. For the RNA-seq dataset (GSE85568), we used the Ensembl gene annotations provided in the GEO entry. These data are summarized in Supplementary Table 2.

Outlier identification

To identify outlier samples within each study, we used Bolstad’s relative log expression (RLE) method [11]. We defined outliers as samples with an interquartile range (IQR) > mean(IQR)+2.5*SD(IQR). This resulted in the exclusion of 17 samples from six studies, as shown in Table 1 in the main text.

Enrichment analyses

We used Gene Set Variation Analysis (GSVA) [12] to transform gene expression data from each study to pathway level scores. Briefly, GSVA works by estimating a gene set score in an unsupervised way by calculating an enrichment statistic for each gene set in each sample. 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 association testing and 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/). We tested gene sets representing biological pathways (BioCarta_2016, KEGG_2016, NCI-Nature_2016, Panther_2016, Reactome_2016, Human_Gene_Atlas, WikiPathways_2016), sets representing regulatory targets (ChEA_2016, ENCODE_Histone_Modifications_2015, ENCODE_TF_ChIP-seq_2015,
Supplementary results on overlap of asthma DEGs with other allergic disease states

We examined whether the top DEGs from the meta-analysis were also associated with other allergic disorders, including allergic rhinitis (AR), atopic dermatitis (AD), and eosinophilic esophagitis (EoE). We identified DEGs for each of these studies from published findings[13–15] and conducted hyper-geometric tests of enrichment. As expected, the overlap between gene sets identified in our meta-analysis with AR was quite strong statistically (hyper-geometric p-value = 6.52 x 10^{-5}), with nine significant genes shared in common, most prominently CST1. Notably, all nine genes in common were modulated in the same direction (Supplementary Table 9). Overlap with AD, while statistically significant, was not impressive in that only two genes were shared in common, one of which was discordant in terms of direction of expression. We did, however, find strong enrichment of asthma DEGs in EoE (p = 4.8 x10^{-10}), though the directionality of gene expression changes was not completely concordant (Supplementary Table 9). Three genes were shared in common between asthma, AR and EoE, namely CLC, CDH26, and HDC.

References


