Transcriptome networks identify mechanisms of viral and nonviral asthma exacerbations in children

Matthew C. Altman [1,2]*, Michelle A. Gill[3], Elizabeth Whalen[2], Denise C. Babineau[4], Baomei Shao[3], Andrew H. Liu[5], Brett Jepson[4], Rebecca S. Gruchalla[3], George T. O'Connor[6], Jacqueline A. Pongracic[7], Carolyn M. Kercsmar[8], Gurjit K. Khurana Hershey[8], Edward M. Zoratti[9], Christine C. Johnson[9], Stephen J. Teach[10], Meyer Kattan[11], Leonard B. Bacharier[12], Avraham Beigelman[12], Steve M. Sigelman[13], Scott Presnell [2], James E. Gern[14], Peter J. Gergen[13], Lisa M. Wheatley[13], Alkis Togias[13], William W. Busse[14] and Daniel J. Jackson[14]

[1]Department of Allergy and Infectious Diseases, University of Washington, Seattle, WA, USA. [2]Systems Immunology Program, Benaroya Research Institute, Seattle, WA, USA. [3]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, USA. [4]Rho, Inc., Chapel Hill, NC, USA. [5]Department of Allergy and Immunology, Children's Hospital Colorado, University of Colorado School of Medicine, Aurora, CO, USA. [6]Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [7]Ann Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA. [8]Cincinnati Children's Hospital, Cincinnati, OH, USA. [9]Henry Ford Health System, Detroit, MI, USA. [10]Children's National Health System, Washington, DC, USA. [11]Columbia University College of Physicians and Surgeons, New York, NY, USA. [12]Division of Allergy, Immunology, and Pulmonary Medicine, Washington University, St. Louis, MO, USA. [13]Division of Allergy, Immunology, and Transplantation, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA. [14]University of Wisconsin School of Medicine and Public Health, Madison, WI, USA. *e-mail: maltman@benaroyaresearch.org
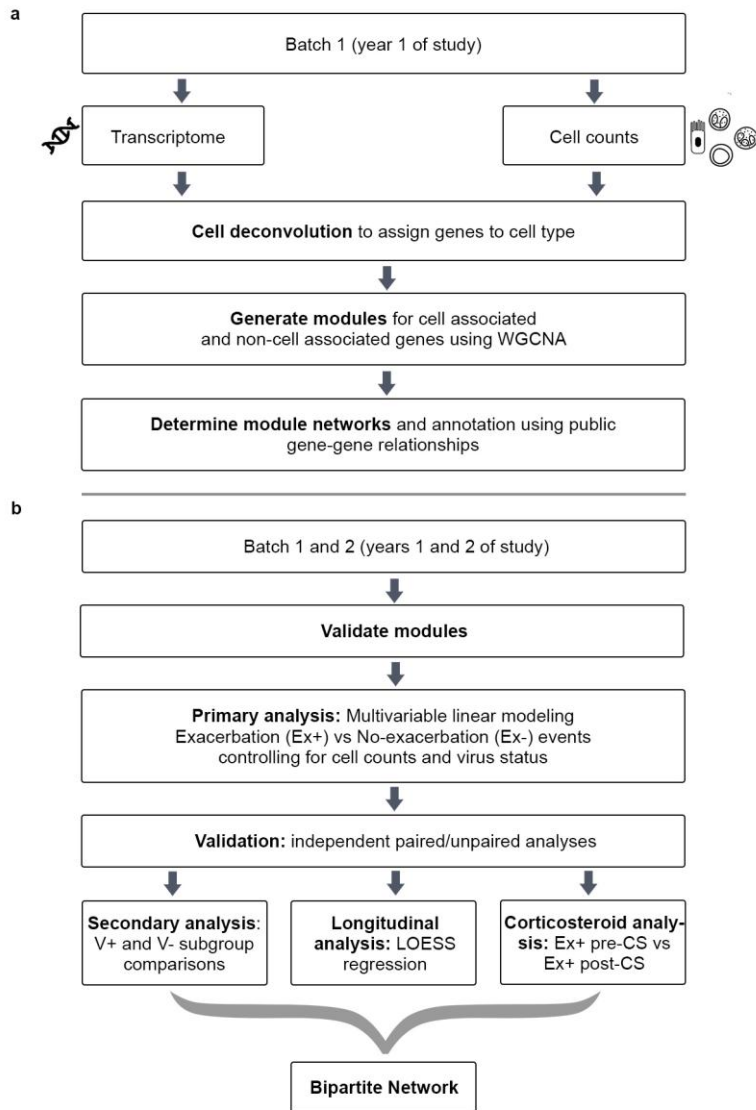
Supplementary Figure 1

Analysis overview

(a) Cell-associated gene coexpression modules were generated using samples collected from the first fall enrollment season, which included 42 participants capturing 59 cold events. Cell deconvolution was performed to assign genes to cell types, and then WGCNA was used to generate coexpression modules. Module networks were determined by assessing known gene-gene interactions and annotated using public databases. (b) These modules were validated as coherent using the full dataset of 106 participants and 154 cold events. The modules were then used for the comparison analyses. The primary analysis compared Ex+ events (47) to Ex- events (107) controlling for cell differentials and presence or absence of a virus-associated with the event (virus status), visit, and library depth with a random effect included for participant. The validation of the primary analysis was done through independent analyses of paired samples (19 participants who had one Ex+ and one Ex- event) and unpaired samples (87 participants who collectively had 28 Ex+ and 88 Ex- events). The secondary analysis compared V+Ex+ (33), V+Ex- (69), V-Ex+ (14), and V-Ex- (38) events. The longitudinal analysis used LOESS regression to estimate the longitudinal dynamics of module expression per day. The corticosteroid analysis compared the expression of modules before and after initiation of systemic corticosteroids in the Ex+ group. All results were merged to generate overview bipartite networks demonstrating the global signals of module expression specific to V+Ex+ events, V-Ex+ events, and the effects of systemic corticosteroids.
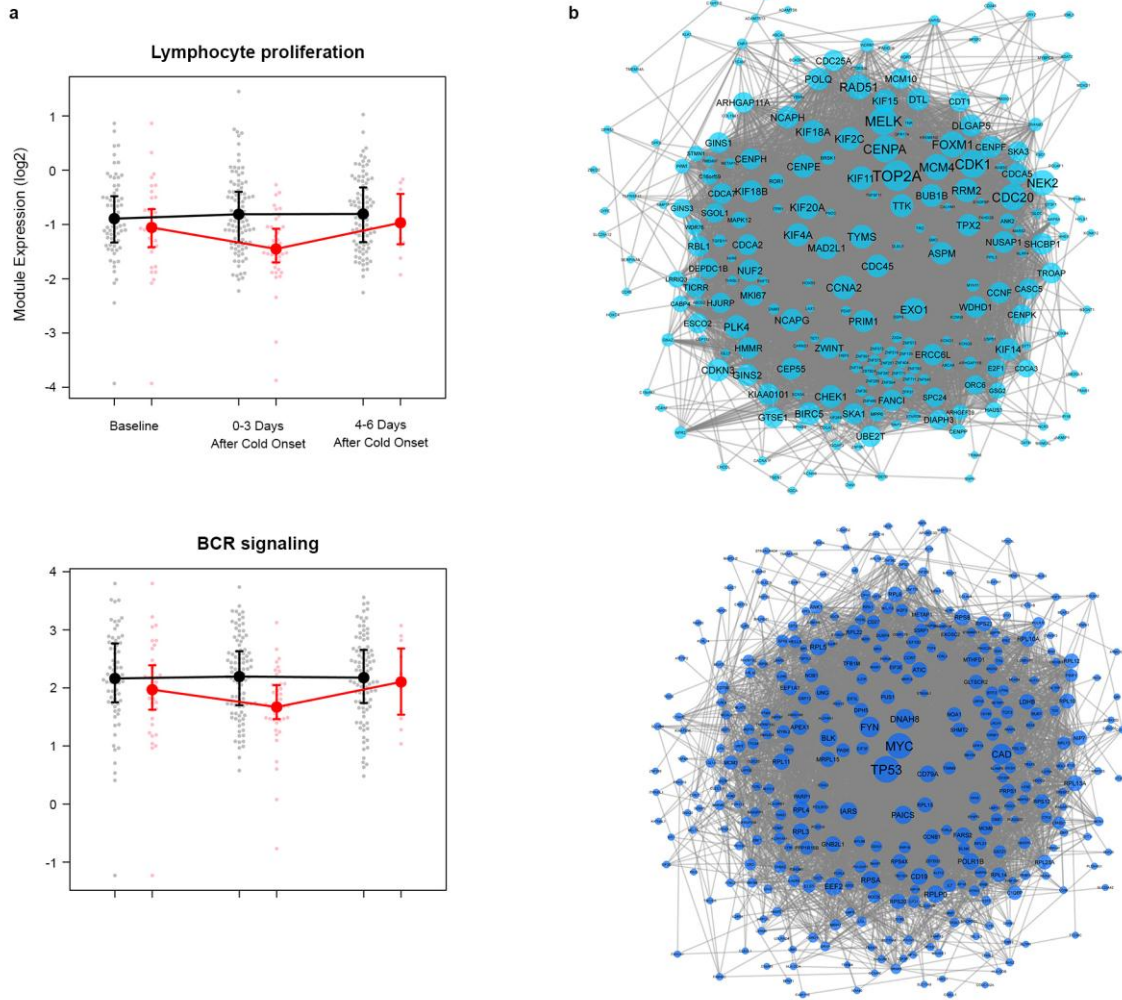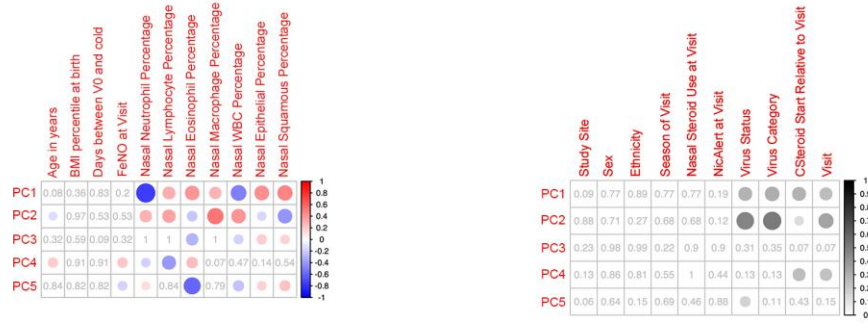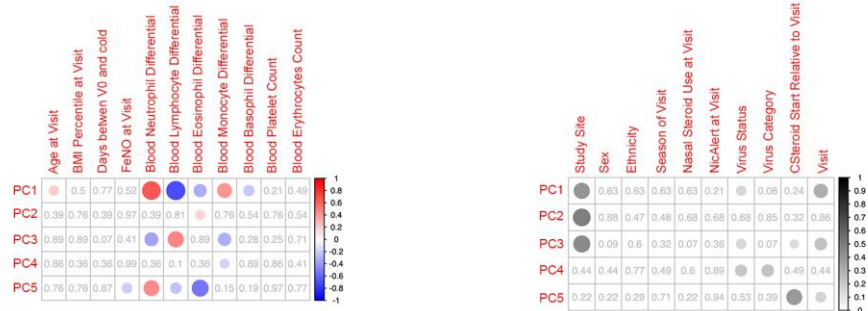
Supplementary Figure 2

Two core exacerbation modules are downregulated in cold events that lead to exacerbation.

(a) The expression levels of the 2 nasal gene expression modules that decreased in Ex+ events compared to Ex- events. FDR values from top to bottom are 1.38E-05, 4.09E-04. Expression levels represent the log base 2 of the geometric mean of the normalized expression of all genes within the module. Shown are group mean values, interquartile ranges, and all data points. Sample sizes are: Ex+ at baseline (n=38); Ex+ at 0-3 days (n=44); Ex+ at 4-6 days (n=11); Ex- at baseline (n=68); Ex- at 0-3 days (n=97); Ex- at 4-6 days (n=95). 18 participants who had one Ex+ and one Ex- event have the same measurement shown in both Ex+ at baseline and Ex- at baseline. All measurements shown are otherwise biologically independent. Comparisons were performed using a weighted linear model and empirical Bayes method, including terms for exacerbation status, cell percentages, presence or absence of virus, visit, and library depth with a random effect included for participant. (b) Gene-gene association networks for each of the modules demonstrate significant interaction networks. Genes are represented as circular nodes, and known gene-gene interactions from STRING as connecting edges. The size of each node is proportional to the number of interactions. STRING enrichment p-values for both networks are <1.0E-16.
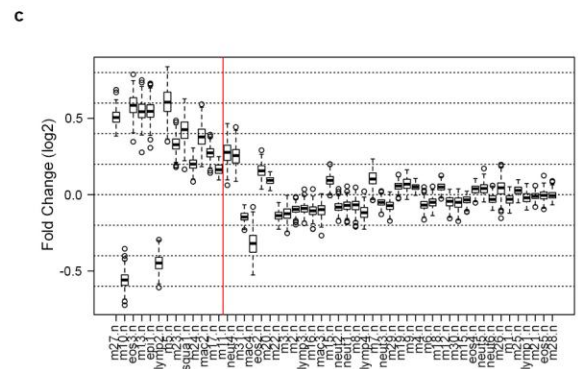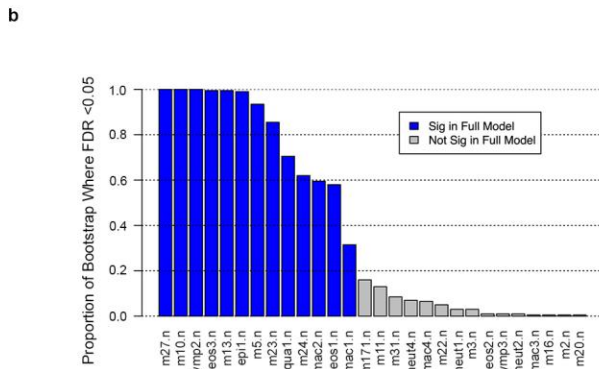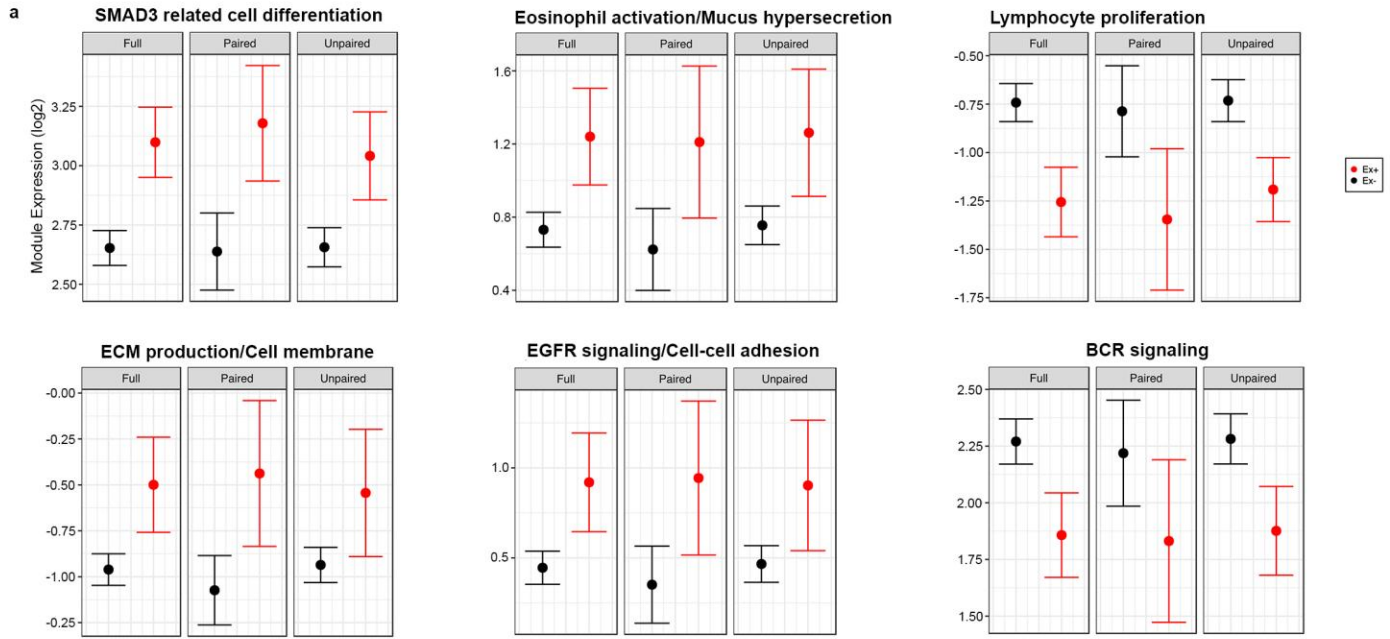
Supplementary Figure 3

Principal component analysis (PCA) shows important sources of variability in the transcriptome.

PCA was performed on the expression data and principal components (PCs) were correlated to clinical and demographic variables. To determine significance, Pearson correlation was calculated for continuous variables and ANOVA was run for categorical variables. Significant relationships (FDR<0.05) between variables and PCs show a circle on the heatmap to indicate magnitude of the relationship, where color represents magnitude and direction of Pearson correlation for continuous variables and darkness of gray represents the magnitude of R2 from ANOVA for categorical variables. Variables that do not have a significant relationship with PCs show the FDR value in their square on the heatmap. (a) Association heatmaps of variables for nasal samples. (b) Association heatmaps of variables for blood samples.

Supplementary Figure 4

Validation analyses demonstrate the core exacerbation modules are reproducible.

(a) Module expression levels for the 6 exacerbation core modules are shown comparing the full cohort (154 total events), the paired cohort (38 total events), and the unpaired cohort (116 total events). The paired cohort included 19 participants who experienced one Ex+ and one Ex- event during the study. The unpaired cohort compared participants who contributed only Ex+ or only Ex- events, and included the other 87 participants who collectively had 28 Ex+ and 88 Ex- events. The two groups were, by definition, independent of one another. The four core modules that were significantly higher in Ex+ events (left-most four plots) have an FDR<0.05 in each comparison. The two modules that were significantly lower in Ex+ events (right-most two plots) have an FDR<0.05 in the full and unpaired cohorts and FDRs of 0.06 (Lymphocyte proliferation) and 0.22 (BCR signaling) in the paired cohort. Expression levels represent the log base 2 of the geometric mean of the normalized expression of all genes within the module. Shown are the group mean and standard error values. (b) A sensitivity analysis demonstrates the proportion of times each module was significant using iterative boot strapping to random subsets to 80% of participants. The first 6 modules are the core modules shown in (a) and were significant in >99% of iterations. The 13 modules in blue are those that were significant in the primary analysis. Modules that were never significant are not shown. Abbreviated alphanumeric module names are listed for brevity. (c) Shown are the associated log2 fold change values for the sensitivity analysis. Calculated fold changes across all iterations are shown as box plots of the medians, interquartile ranges (boxes), and 1.5 times the interquartile ranges (whiskers) with outliers shown as points. The red line indicates the 13 modules significant in the primary analysis. Abbreviated alphanumeric module names are used for brevity.

Supplementary Figure 5

Core exacerbation modules are common to exacerbation events with or without virus present, while additional distinct patterns of module expression are unique to V+Ex+ and V-Ex+ events.
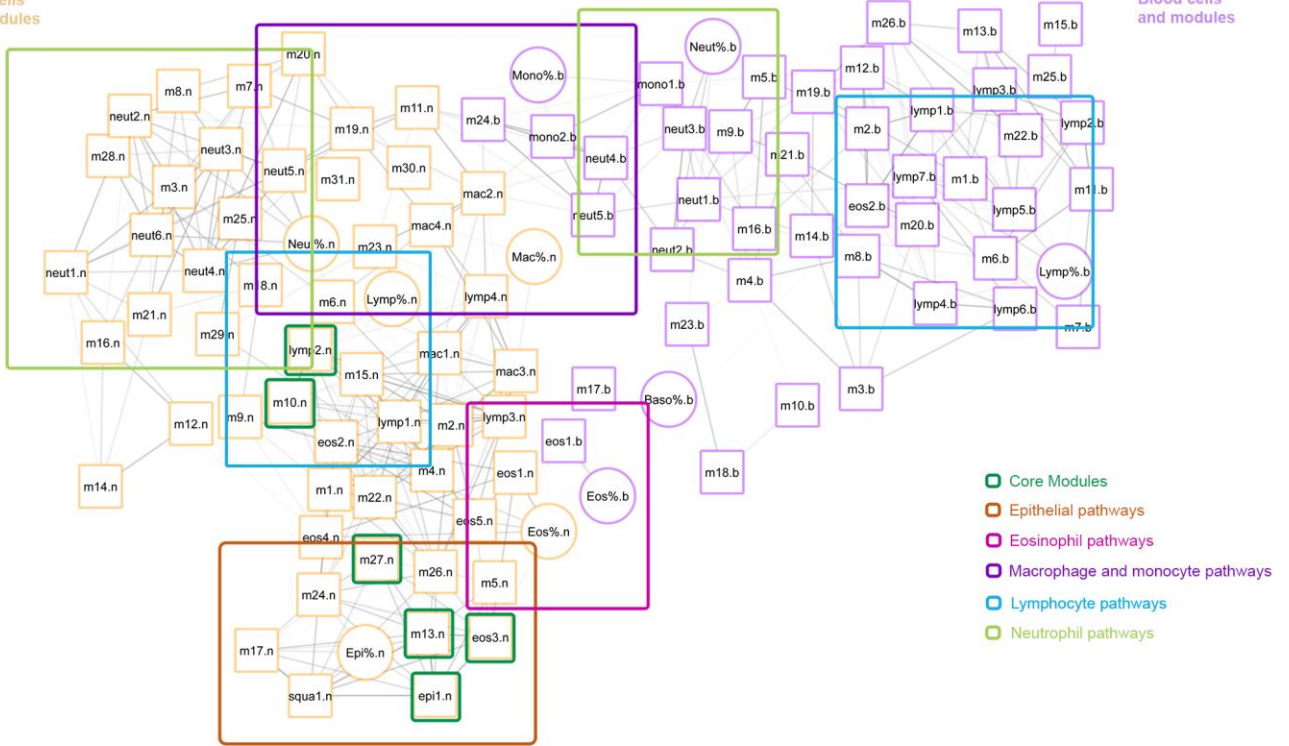
(a) The core exacerbation modules show a similar degree of altered expression in both V+Ex+ (n=33) and V-Ex+ (n=14) events and are significant in each subgroup comparison. (b) V+Ex+ events demonstrate specific upregulation in 9 modules that represent numerous inflammatory pathways and specific downregulation in 2 modules. (c) V-Ex+ events demonstrate specific upregulation of 3 squamous

cell associated modules. Expression levels represent the geometric mean of the normalized expression of all genes within the module. Shown are subgroup mean and standard error values. All modules shown have an ANOVA FDR<0.05. FC and FDR values are listed in Table 2. This analysis included 247 unique nasal samples and 256 unique blood samples from 106 individuals who in total had 33 V+Ex+, 14 V-Ex+, 69 V+Ex- and 38 V-Ex- events.
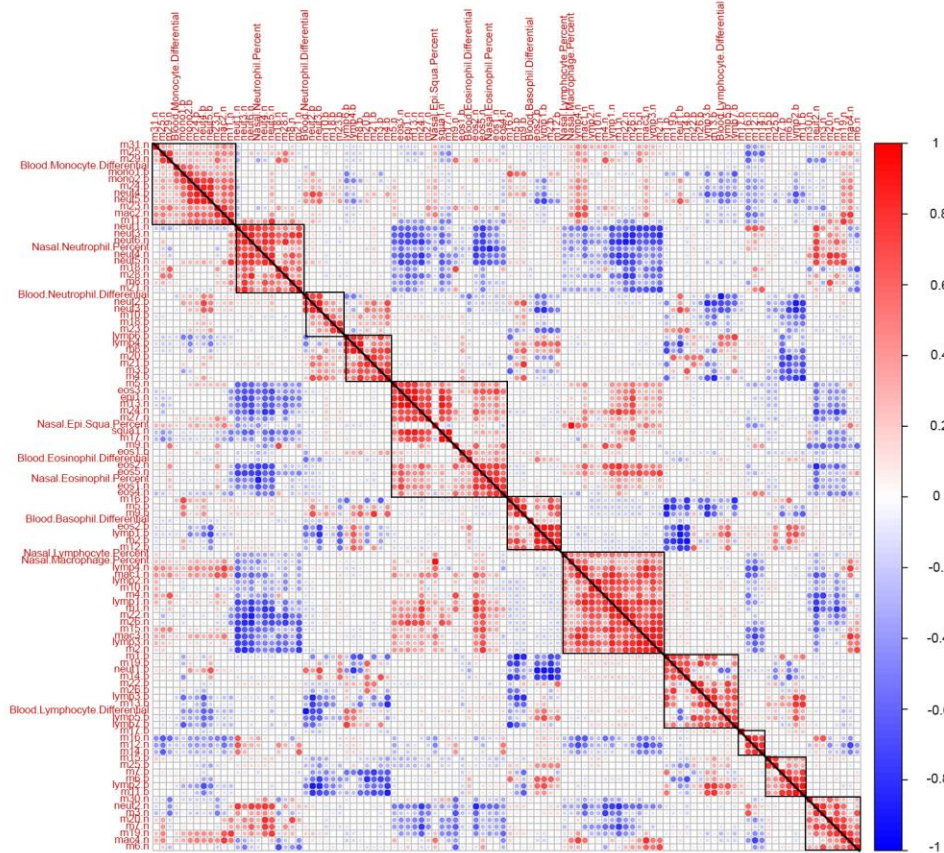
a

Nasal cells
and modules

Blood cells
and modules

Core Modules
Epithelial pathways
Eosinophil pathways
Macrophage and monocyte pathways
Lymphocyte pathways
Neutrophil pathways

b

Supplementary Figure 6

Expression profiles define clusters of biological functions.

(a) A bipartite network demonstrates the coassociations among all nasal (light orange) and blood (light purple) modules (squares) and cell percentages (circles). Edges represent significant positive Pearson correlations >0.5 (FDR<0.05) and darker edges indicate higher correlations. Nodes are clustered according to their interconnectedness. Because respiratory epithelium and squamous cells were highly correlated, they were combined for this analysis. Several co-clustered networks are outlined in colors. Nasal epithelial (brown outline), nasal and blood eosinophil (pink outline) cells and pathways cluster towards the bottom of the network. These clusters include the upregulated core modules (Green squares) including pathways of multiple epithelial functions, eosinophil activation, and Th2 inflammation. Nasal macrophage and blood monocyte cells and modules cluster with one another (dark purple outline). This cluster includes blood and nasal interferon responses, the heat shock protein response, airway macrophage pathways, and blood monocyte pathways. Nasal lymphocyte pathways are not clustered with blood lymphocytes pathways (blue). Two core modules fall within the nasal lymphocyte cluster (green squares). Nasal neutrophil and blood neutrophil pathways do not cluster with one another (light green). (b) The heatmap shows the pairwise Pearson correlation coefficients between all nasal and blood module expression levels and nasal and blood cell percentages. Red indicates positive correlation and blue indicates negative correlation. Clusters represent coassociated modules and cell types. Abbreviated alphanumeric module names are used for brevity.
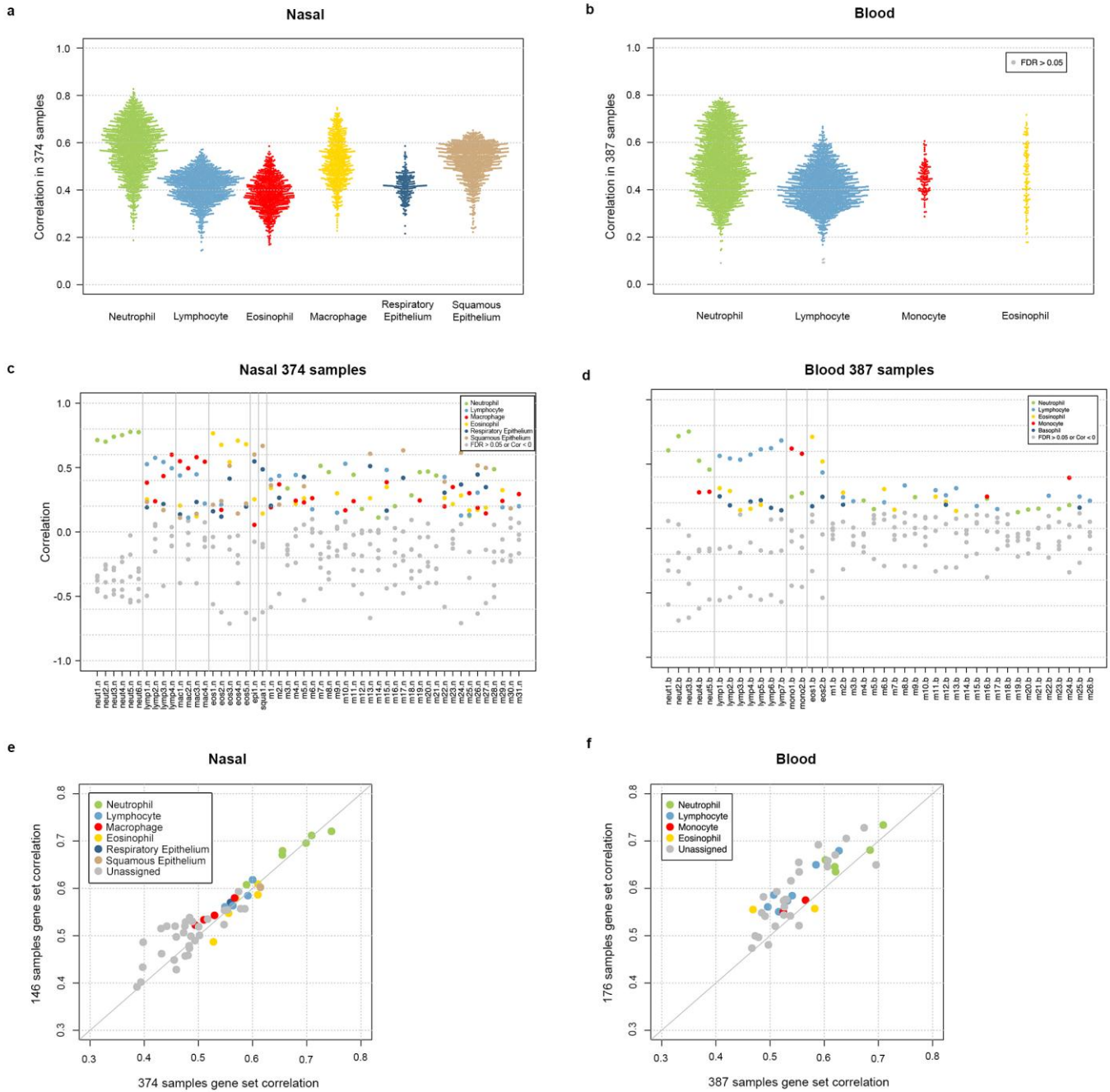
Supplementary Figure 7

Module validation demonstrates that module-cell associations and module coherence are reproducible.

(a) In the full set of 374 nasal samples, all genes that had been associated to a specific cell type during module generation, had statistically significant positive Pearson correlations with the same cell type (13,672 genes). (b) In the full set of 387 blood samples, all but 5 genes that had been associated to a specific cell type during module generation, had statistically significant positive Pearson correlations with the same cell type (13,311/13,316 genes). (c) Correlation of nasal module expression levels from the full set of 374 samples with cell differentials was assessed to determine a module's final cell assignment. 27/52 nasal modules were assigned to one or more cell types. (d) Correlation of blood module expression levels from the full set of 387 samples with cell differentials to finalize a module's cell assignment. 16/42 blood modules were assigned to one cell type. (e) The average Pearson correlation of genes assigned to each nasal module during module construction (146 samples) versus the average Pearson correlation of genes assigned to the

same module within the full set of 374 samples. (f) The average Pearson correlation of genes assigned to each blood module during module construction (176 samples) versus the average Pearson correlation of genes assigned to the same module within the full set of 387 samples.