nature immunology

RESOURCE

https://doi.org/10.1038/s41590-019-0347-8

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]

Respiratory infections are common precursors to asthma exacerbations in children, but molecular immune responses that determine whether and how an infection causes an exacerbation are poorly understood. By using systems-scale network analysis, we identify repertoires of cellular transcriptional pathways that lead to and underlie distinct patterns of asthma exacerbation. Specifically, in both virus-associated and nonviral exacerbations, we demonstrate a set of core exacerbation modules, among which epithelial-associated SMAD3 signaling is upregulated and lymphocyte response pathways are downregulated early in exacerbation, followed by later upregulation of effector pathways including epidermal growth factor receptor signaling, extracellular matrix production, mucus hypersecretion, and eosinophil activation. We show an additional set of multiple inflammatory cell pathways involved in virus-associated exacerbations, in contrast to squamous cell pathways associated with nonviral exacerbations. Our work introduces an in vivo molecular platform to investigate, in a clinical setting, both the mechanisms of disease pathogenesis and therapeutic targets to modify exacerbations.

Exacerbations are the primary cause of morbidity and mortality in children with asthma and occur despite current treatments. It has been established that the majority of exacerbations are provoked by viral respiratory infections, most notably rhinoviruses (RVs); however, exacerbations also occur in the absence of infections[1,2]. Furthermore, even in children with severe asthma, the majority of respiratory infections do not progress to exacerbation[3]. The fundamental question of why some but not all colds lead to exacerbations has not been fully answered, and the similarities and differences between the mechanism(s) of virus-associated and nonviral exacerbations are unknown. Identifying the molecular pathways underlying progression from colds to exacerbations has the potential to improve understanding of the mechanisms of these events and, from this information, to target therapeutic strategies more precisely and effectively.

Beneficial effects in exacerbation prevention have been seen with type 2 and eosinophil-modulating monoclonal antibodies, supporting the importance of related pathways[4–6]. However, these therapies reduce, rather than eliminate, exacerbations and do not correct the physiological abnormalities of asthma, thereby confirming the contribution of additional factors to the genesis of exacerbations[7]. Respiratory epithelial inflammation, in particular IL-33 signaling, has been demonstrated to have an important role in RV-induced exacerbations[8]. Several additional immune pathways have been linked to asthma exacerbations, generally through targeted investigations[9]. Current evidence has yet to fully establish how multiple interacting molecular pathways may lead to asthma exacerbations.

Systems-scale data collection and network analysis facilitates unbiased mechanistic disease research[10]. This approach has added novel information to the understanding of asthma[11–14]; however, previous asthma 'omics' studies have not been designed to determine molecular networks specific to asthma exacerbations. We have therefore performed the first study in asthma that uses transcriptome network analysis in a case–control, longitudinal study design. In our study, we prospectively evaluated a cohort of children with high disease burden. We combined nasal virus PCR, nasal and peripheral blood cell differentials, and nasal and blood RNA-seq for serial sample collections at baseline and longitudinally during episodes when participants reported cold symptoms to determine the key molecular pathways associated with subsequent asthma exacerbations. We combined cell deconvolution, weighted gene correlation network analysis (WGCNA), and linear mixed-effects

[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

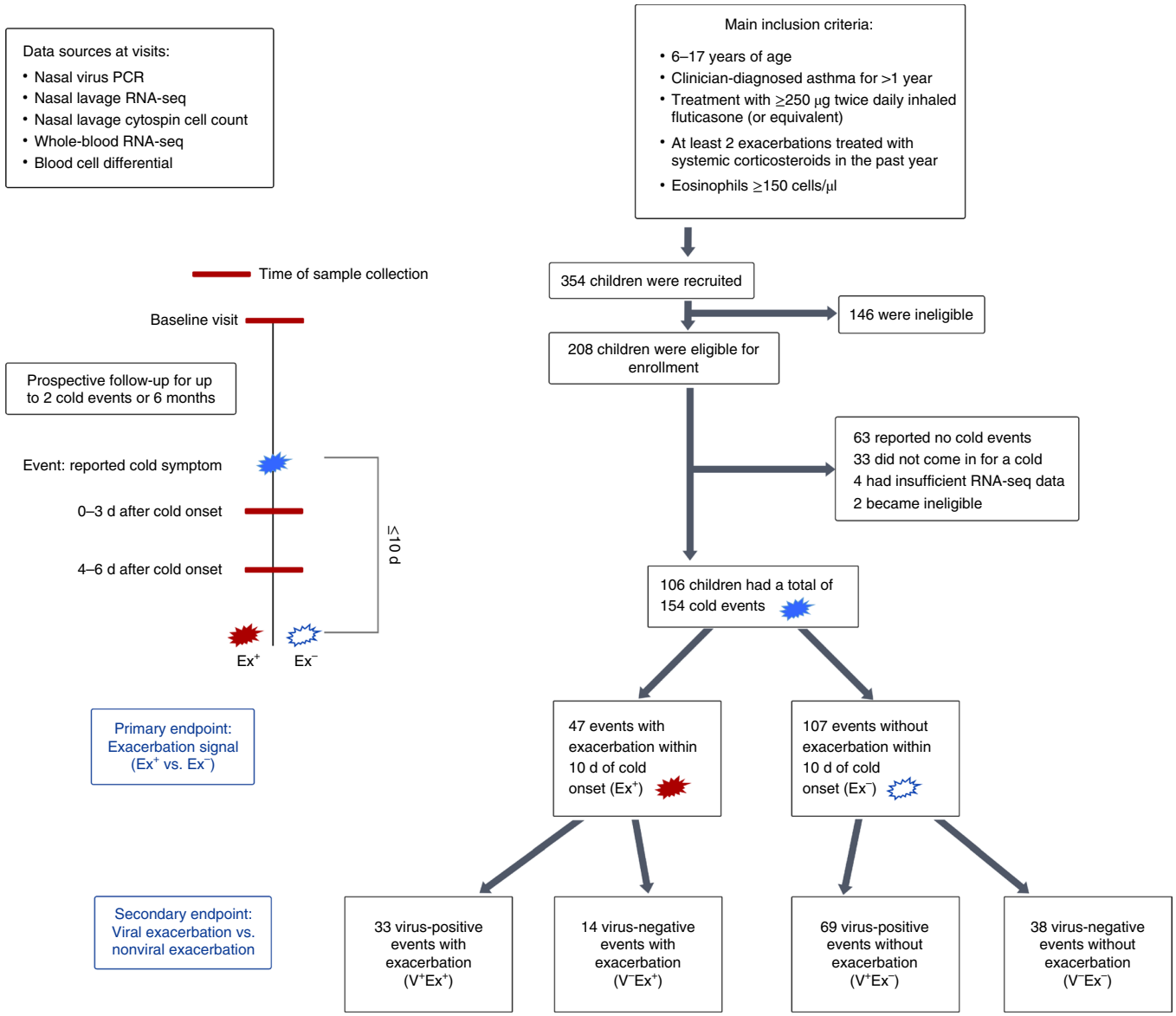


Fig. 1 | Study design and primary and secondary endpoints. 208 children with exacerbation-prone asthma were enrolled according to the inclusion criteria. All participants had baseline samples collected and were prospectively monitored for the onset of cold symptoms (events) during a 6-month period. 106 of the 208 participants came in during one or more events and had sufficient samples collected for analysis. Events were grouped as exacerbations (Ex^+) or not exacerbations (Ex^-), depending on whether the event led to clinical symptoms that resulted in systemic corticosteroid use within 10 d of event onset or resolved without treatment with systemic corticosteroids. The primary endpoint compared Ex^+ events ($n=47$) to Ex^- events ($n=107$). On the basis of nasal virus PCR results, events were further classified as virus associated (V^+) or nonviral (V^-). The secondary endpoint compared V^+Ex^+ ($n=33$), V^-Ex^+ ($n=14$), V^+Ex^- ($n=69$), and V^-Ex^- ($n=38$) events.

modeling to establish the repertoires of cellular transcriptional pathways underpinning asthma exacerbations, including distinct pathways in virus-associated and nonviral exacerbations, the kinetics of these pathways leading to exacerbations, predictors of exacerbation risk, and the responses to systemic corticosteroid therapy. Our results constitute an unbiased, detailed, and sensitive systems-level mechanistic summary that substantially advances understanding of the immunopathogenesis of asthma exacerbations. Moreover, these methods provide a framework for in vivo investigation of mechanisms of therapeutic intervention.

Results

Characterization of colds and exacerbations. We enrolled 208 children 6–17 years old with a daily inhaled corticosteroid (ICS) requirement of at least 500 μg of fluticasone (or the equivalent), a history of at least two exacerbations treated with systemic corticosteroids and/or hospitalization in the previous 12 months, and a blood eosinophil concentration of ≥150 cells/mm^3. We collected nasal lavage and peripheral blood samples at enrollment after a period without cold symptoms or an asthma exacerbation (baseline) and again after the onset of a respiratory illness characterized by common cold symptoms (termed 'events'), including two samples within 6 d of cold symptom onset per event (Fig. 1). Each participant was followed for two events or 6 months.

The primary aim of the study was to identify patterns of gene expression induced during events that progressed to asthma exacerbations (exacerbation, Ex^+), defined by clinical symptoms that resulted in systemic corticosteroid use within 10 d of cold

Table 1 | Population demographics and clinical characteristics

Characteristic		Baseline ($n=106$)	Ex$^+$ events ($n=47$)	Ex$^-$ events ($n=107$)	Ex$^+$ vs. Ex$^-$ *P* value
Age (years) at baseline		11 (8, 13)	12 (9, 14)	11 (8, 13)	0.999
BMI percentile		86.65 (63, 97.8)	80.2 (45.1, 97.4)	86.85 (67.3, 98.3)	0.698a
Sex	Female	49.1% (52)	51.1% (24)	49.5% (53)	0.903
Race	Black (non-Hispanic)	53.8% (57)	61.7% (29)	49.5% (53)	0.396
	White (non-Hispanic)	2.8% (3)	4.3% (2)	3.7% (4)	
	Hispanic	34.9% (37)	31.9% (15)	36.4% (39)	
	Other/mixed	7.5% (8)	2.1% (1)	8.4% (9)	
	Unknown or not reported	0.9% (1)	0.0% (0)	1.9% (2)	
Site	Boston	13.2% (14)	10.6% (5)	15.9% (17)	0.959
	Chicago	6.6% (7)	6.4% (3)	4.7% (5)	
	Dallas	7.5% (8)	2.1% (1)	8.4% (9)	
	Denver	12.3% (13)	23.4% (11)	8.4% (9)	
	Detroit	13.2% (14)	14.9% (7)	13.1% (14)	
	New York	22.6% (24)	17.0% (8)	23.4% (25)	
	Washington DC	5.7% (6)	10.6% (5)	3.7% (4)	
	Cincinnati	6.6% (7)	6.4% (3)	7.5% (8)	
	St. Louis	12.3% (13)	8.5% (4)	15.0% (16)	
FEV_1 predicted (%) at visit a		92 (82.3, 100.9)	72.1 (63.8, 84.2)	93.5 (83.6, 103)	4.70×10^{-6}
FEV_1 predicted (%) at visit b			77.6 (64.4, 85.6)	92.7 (85.2, 105.2)	7.11×10^{-8}
FEV_1 % change from baseline at visit a			−10.3 (−24.1, 3.4)	1.6 (−9.2, 8.9)	3.65×10^{-4}
FEV_1 % change from baseline at visit b			−5 (−24.1, 3.9)	0.5 (−7, 8.6)	2.99×10^{-3}
FEV_1/FVC change from baseline at visit a			−0.06 (−0.12, 0.01)	0.01 (−0.04, 0.06)	7.18×10^{-5}
FEV_1/FVC change from baseline at visit b			−0.02 (−0.06, 0.02)	0.01 (−0.03, 0.05)	2.61×10^{-2}
FeNO at visit a		33 (15.5, 57)	38.5 (17, 76)	33 (15, 68.5)	0.710a
FeNO change from baseline at visit a			3.5 (−11, 20)	3.5 (−9, 19.5)	0.221
Cold severity by WURSS AUC over 10 d after cold			43 (31, 68)	32.5 (18, 56.5)	1.06×10^{-2}
Nasal corticosteroid use within 24 h of visit a	Yes	10.4% (11)	17.0% (8)	15.9% (17)	0.873
Days between cold onset and corticosteroid use			2 (2, 5)		
Season of cold	Fall (Sep.–Nov.)		31.9% (15)	23.4% (25)	0.169b
	Winter (Dec.–Feb.)		31.9% (15)	37.4% (40)	
	Spring (Mar.–May)		19.1% (9)	30.8% (33)	
	Summer (Jun.–Aug.)		17.0% (8)	8.4% (9)	
Days between cold onset and visit a			2 (1, 2)	2 (1, 2)	0.417
Days between cold onset and visit b			5 (4, 6)	5 (4, 6)	0.909
Viral cold type	Rhinovirus A	4.7% (5)	23.4% (11)	13.1% (14)	0.127
	Rhinovirus C	5.7% (6)	19.1% (9)	19.6% (21)	0.984
	Rhinovirus B	6.6% (7)	2.1% (1)	7.5% (8)	0.225
	Enterovirus	0.0% (0)	2.1% (1)	0.0% (0)	0.307[b]
	Rhinovirus/enterovirus (non-specific)	0.0% (0)	0.0% (0)	1.9% (2)	1.00[b]
	Adenovirus	2.8% (3)	0.0% (0)	1.9% (2)	1.00[b]
	Bocavirus	2.8% (3)	2.1% (1)	6.5% (7)	0.334
	Respiratory syncytial virus (A or B)	2.8% (3)	4.3% (2)	2.8% (3)	0.652

Continued

Table 1 | Population demographics and clinical characteristics (continued)

Characteristic		Baseline ($n=106$)	Ex⁺ events ($n=47$)	Ex⁻ events ($n=107$)	Ex⁺ vs. Ex⁻ *P* value
	Coronavirus	4.7% (5)	8.5% (4)	10.3% (11)	0.749
	Parainfluenza virus	1.9% (2)	8.5% (4)	3.7% (4)	0.245
	Metapneumovirus	1.9% (2)	4.3% (2)	2.8% (3)	0.652
	Influenza B	0.0% (0)	0.0% (0)	1.9% (2)	1.00[b]
	Multiple viruses	5.7% (6)	4.3% (2)	7.5% (8)	0.465
	No virus detected	69.8% (74)	29.8% (14)	34.6% (37)	0.476
	Not collected	1.9% (2)	0.0% (0)	0.9% (1)	1.00[b]
Urine cotinine level	Level 0 (0–10 ng/ml)	27.4% (29)	23.4% (11)	16.8% (18)	0.160[b]
	Level 1 (10–30 ng/ml)	41.5% (44)	36.2% (17)	53.3% (57)	
	Level 2 (30–100 ng/ml)	20.8% (22)	29.8% (14)	15.0% (16)	
	Level 3 (100–200 ng/ml)	9.4% (10)	8.5% (4)	10.3% (11)	
	Level 4 (200–500 ng/ml)	0.9% (1)	0.0% (0)	0.9% (1)	
	Level 5 (500–1,000 ng/ml)	0.0% (0)	0.0% (0)	0.9% (1)	
Blood neutrophil differential (%)		45 (37, 53)	50 (42, 61)	51 (42.6, 60)	0.997
Blood lymphocyte differential (%)		39.3 (32.9, 46.1)	31.2 (23.9, 38.1)	32.6 (27, 40)	0.257
Blood monocyte differential (%)		7.1 (6.1, 9.1)	7.8 (6, 10.1)	7.3 (6.4, 9.5)	0.331
Blood eosinophil differential (%)		6 (4.3, 8.8)	7 (5, 9.4)	5.2 (3.4, 9.2)	0.269
Blood basophil differential (%)		0.8 (0.6, 1)	0.6 (0.4, 1)	0.6 (0.2, 1)	0.836
Nasal neutrophil differential (%)		67.7 (36.6, 80.8)	75.2 (49.7, 89)	77.8 (51.6, 91.4)	0.803
Nasal lymphocyte differential (%)		1.2 (0.5, 2.6)	1.3 (0.6, 3.1)	1.4 (0.6, 3.4)	0.209
Nasal macrophage differential (%)		3.2 (1.7, 5.3)	3.7 (2.4, 8.6)	3.8 (2.2, 6.6)	0.577
Nasal eosinophil differential (%)		4 (0.8, 11.5)	3.2 (1.3, 8.1)	2.4 (0.2, 9.6)	0.916
Nasal epithelium differential (%)		2 (0, 4)	1 (0, 4.5)	1 (0, 3)	0.728
Nasal squamous differential (%)		13 (3, 42)	3 (1, 10)	4 (2, 17)	0.206

106 participants came in during one or more cold illnesses contributing in total 154 events. 47 events were classified as exacerbation associated (Ex⁺) and 107 were classified as not exacerbation associated (Ex⁻). Ex⁺ events were characterized by significantly lower pulmonary function measurements during the illness and higher cold severity symptoms in the 10 d following cold onset. WURSS AUC, Wisconsin Upper Respiratory Symptom Survey area under the curve. There were no differences between the event groups with respect to age, body mass index (BMI), sex, race, enrollment city (site), fractional exhaled nitric oxide (FeNO), season of event, virus type, urine cotinine level, or blood or nasal cell differentials. Percent (count) is displayed for categorical variables. Median (first quartile, third quartile) is displayed for continuous variables. For participants with two events, data for both colds are included in the table. Unless otherwise specified, all comparisons were performed by using a generalized linear mixed model with a random effect for participant to account for correlation between values from the same participant; a multinomial distribution was used for categorical variables and a normal distribution was used for continuous variables. [a]Analyzed by log-normal distribution. [b]Analyzed by two-sided Fisher's exact test.

symptom onset, versus those events that resolved without treatment with systemic corticosteroids (no exacerbation, Ex⁻). Of the 154 total events captured in 106 children, 47 were Ex⁺ events and 107 were Ex⁻ events. All Ex⁺ events had one or two samples collected prior to initiation of systemic corticosteroids, and these samples were used for the primary analysis. When comparing the Ex⁺ and Ex⁻ events, there were no differences in demographic variables or in nasal or blood cell differentials (Table 1). Virus PCR and partial sequencing were used to identify events associated with a respiratory virus (V⁺) or virus-negative (V⁻) events; 66.2% of events had a detectable respiratory virus, and, of these, 53.9% were positive for RV-A or RV-C. However, there were no differences in virus frequency or type when comparing Ex⁺ and Ex⁻ events. Pulmonary functions were assessed at baseline and during events. Only Ex⁺ events were associated with significant decline in forced expiratory volume in 1 s (FEV_1) predicted (%) and FEV_1 to forced vital capacity (FEV_1/FVC) ratio, demonstrating a change in pathophysiological measures of lung function associated with the standardized clinical definition of asthma exacerbation used in the study (Fig. 2).

These results confirm the high rate of asthma exacerbations associated with cold symptoms in this population and the corresponding pathophysiological changes associated with exacerbations.

Modular immune responses underlying asthma exacerbations. From nasal and blood samples, we combined measured cell percentages with RNA-seq data and used cell deconvolution and WGCNA to construct and validate a repertoire of 94 distinct gene expression modules representing a diverse array of biological functions (Supplementary Fig. 1 and Supplementary Table 1). Gene expression levels measured in nasal samples were associated with percentages of neutrophils, lymphocytes, macrophages, eosinophils, respiratory epithelial cells (termed 'epithelium'), and/or squamous epithelial cells (termed 'squamous') in nasal samples. Gene expression levels measured in peripheral blood were associated with percentages of neutrophils, lymphocytes, monocytes, eosinophils, and/or basophils in peripheral blood. By applying this modular framework, we compared Ex⁺ and Ex⁻ events to identify immune pathways underlying asthma exacerbations. Expression patterns in 13 nasal modules were statistically significantly associated with exacerbations at false-discovery rate (FDR) < 0.05, 11 of which had increased expression in Ex⁺ events and 2 of which had decreased expression (Table 2); 6 of these modules had FDR < 0.001 (Fig. 3a and Supplementary Fig. 2a). In contrast, there were no significant differences in expression for the blood modules. The comparison was adjusted for cell percentages and the presence or absence

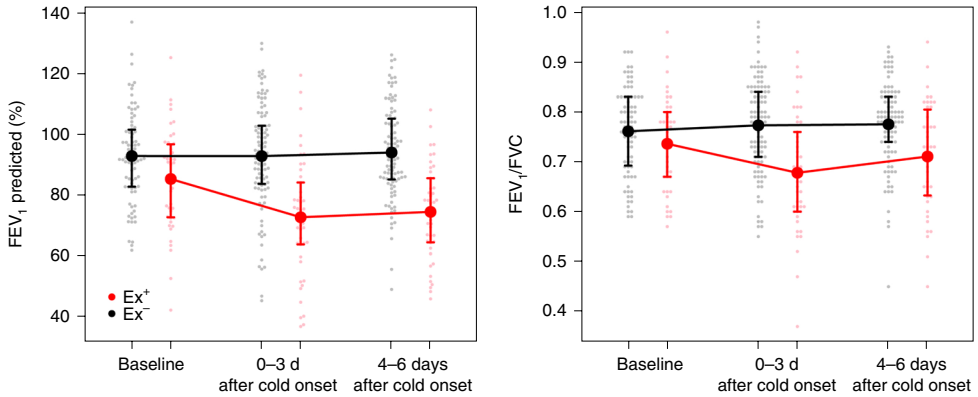


Fig. 2 | Pulmonary function declines during colds that lead to an exacerbation. Pulmonary function measured by FEV_1 predicted (%) and FEV_1/FVC ratio was similar at baseline in participants in the Ex^+ (red) and Ex^- (black) event groups and showed a significant decline during reported cold symptoms in Ex^+ events but not in Ex^- events. Data are shown as the mean, interquartile range, and all data points. *P* values for comparison of the Ex^+ and Ex^- groups are as follows: FEV_1 predicted (%) at 0–3 d, $P=4.7\times10^{-6}$; FEV_1 predicted (%) at 4–6 d, $P=7.1\times10^{-8}$; FEV_1/FVC at 0–3 d, $P=2.3\times10^{-5}$; FEV_1/FVC at 4–6 d, $P=4.6\times10^{-3}$. The number of measurements represented for each group and time point is as follows: Ex^+ at baseline, $n=43$; Ex^+ at 0–3 d, $n=45$; Ex^+ at 4–6 d, $n=42$; Ex^- at baseline, $n=82$; Ex^- at 0–3 d, $n=103$; Ex^- at 4–6 d, $n=97$. 19 participants who had one Ex^+ and one Ex^- event have the same measurement shown for Ex^+ at baseline and Ex^- at baseline. All measurements shown are otherwise biologically independent. Comparisons were performed by using a generalized linear mixed model with a random effect for participant to account for correlation between values from the same participant.

of virus during cold symptoms; these variables were the primary sources of variability in module expression (Supplementary Fig. 3) but were not different between the Ex^+ and Ex^- groups. Our results therefore suggest that the observed differences are predominantly due to changes in gene expression in specific cell types as opposed to changes in cell population composition. Each of the differentially expressed modules represents a significant protein–protein interaction network (enrichment $P<0.001$) that includes important known asthma-associated genes but also demonstrates a much broader functional network (Fig. 3b and Supplementary Fig. 2b).

We annotated the most significant exacerbation-upregulated module as the epithelial-associated 'SMAD3-related cell differentiation' module, which was associated with epithelium and squamous percentages. It contains 46 genes that form a network centered around the transcription factor *SMAD3* and includes other transcription factors and signaling molecules induced by transforming growth factor (TGF)-β1 signaling, including *KLF4*, *FZD1*, and *FZD5*, as well as genes involved in wound healing and epithelial differentiation. The second most significant upregulated module was annotated as 'eosinophil activation/mucus hypersecretion'. It contains 69 genes and was associated with eosinophil, epithelium, and squamous percentages; it includes the key epithelial mucin associated with hypersecretion, *MUC5AC*[15], the respiratory epithelial proliferation marker *KRT8*[16], the eosinophil surface activation/survival molecule *CD9*[17], and the eosinophil transcription factor *IKZF2*[18]. *FGFR2*, a receptor implicated in dysregulation of epithelial–mesenchymal signaling in asthma[19], represents the central-most gene within this module network. The third most significant upregulated module, 'extracellular matrix (ECM) production/cell membrane', contains 209 genes and includes the asthma-associated collagen *COL1A1*[20], the membrane molecules *CAV1* and *CFTR*, and the inducible nitric oxide synthase *NOS2*[21]. The core gene of this network is *EPHA1*, a tyrosine kinase linked to pulmonary inflammation[22]. This module was associated with epithelium and squamous percentages. The fourth most significant upregulated module, 'epidermal growth factor receptor (EGFR) signaling/cell–cell adhesion', contains 240 genes and represents an important pathway associated with lung fibrosis[23] and mucus secretion[24]. It forms a densely clustered network centered around *EGFR* and its homolog receptor *ERBB2* and includes associated signaling molecules; it was associated with epithelium and squamous percentages.

The most significant module downregulated in exacerbation, 'lymphocyte proliferation', was associated with nasal lymphocyte percentage and is a network of cell cycle genes. The other downregulated module was annotated as 'B cell receptor (BCR) signaling' and forms a network of both BCR signaling genes and genes related to mRNA translation; it was associated with lymphocyte percentage.

These results demonstrate upregulation of multiple respiratory epithelium–, squamous-, and eosinophil-associated responses and downregulation of lymphocyte-associated responses as molecular mechanisms in events that lead to asthma exacerbations as compared to those that resolve without exacerbation.

Validation of modular signatures associated with exacerbations. To cross-validate these findings, the cohort was split into prespecified paired and unpaired analyses. The paired analysis included 19 participants who experienced one Ex^+ and one Ex^- event during the study who thus had a sample that could serve as their own control. The unpaired analysis 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. By definition, these validation groups were mutually exclusive.

The paired analysis demonstrated significant differences in expression patterns for seven nasal modules and the unpaired analysis demonstrated such differences for six nasal modules (FDR < 0.05). The top four most statistically significant modules with increased expression in Ex^+ events were the same in both comparisons (FDR < 0.05) and were the same modules observed in the full-cohort comparison discussed in the previous section. These modules showed a statistically similar level of difference in expression in each independent comparison. Similarly, the top two modules with decreased expression in Ex^+ events were the same in each comparison and the same modules found in the full-cohort comparison and discussed in the previous section. In the larger unpaired cohort, each of these modules met a stringent significance threshold of FDR < 0.05; in the smaller paired cohort, they were each borderline significant with FDR values of 0.06 and 0.22 (Supplementary Fig. 4a).

To further test the robustness of our findings, we performed a bootstrap sensitivity analysis through iterative subsetting of the full population. The same six modules were significant with

Table 2 | Results of the primary and secondary analyses

	Summary annotation	Sample type	Cell association	Number of genes	Primary analysis			Secondary analysis				
					Full cohort—exacerbation modules: Ex^+ vs. EX^-			Virus-associated exacerbation: V^+Ex^+ vs. V^+Ex^-		Nonviral exacerbation: V^-Ex^+ vs. V^-Ex^-		
					Fold change	*P*	FDR	Fold change	Two-group *P*	Fold change	Two-group *P*	Four-group ANOVA FDR
Core modules	SMAD3-related cell differentiation	Nasal	Epithelium	46	1.43	3.90×10^{-9}	2.03×10^{-7}	1.46	7.93×10^{-8}	1.34	3.75×10^{-3}	1.37×10^{-6}
	Eosinophil activation/mucus hypersecretion	Nasal	Eosinophil, epithelium	69	1.50	1.19×10^{-6}	1.92×10^{-5}	1.50	3.99×10^{-5}	1.50	4.32×10^{-3}	3.93×10^{-6}
	Extracellular matrix production/cell membrane	Nasal	Epithelium	209	1.46	1.47×10^{-6}	1.92×10^{-5}	1.49	1.93×10^{-5}	1.40	1.19×10^{-2}	1.78×10^{-5}
	Epidermal growth factor receptor signaling/cell–cell adhesion	Nasal	Epithelium	240	1.46	5.37×10^{-6}	5.59×10^{-5}	1.47	1.09×10^{-4}	1.46	8.62×10^{-3}	7.71×10^{-5}
	Lymphocyte proliferation	Nasal	Lymphocyte	266	0.68	5.32×10^{-7}	1.38×10^{-5}	0.65	2.11×10^{-6}	0.75	2.97×10^{-2}	5.99×10^{-5}
	B cell receptor signaling	Nasal	Lymphocyte	378	0.74	4.72×10^{-5}	4.09×10^{-4}	0.72	2.12×10^{-4}	0.78	4.98×10^{-2}	2.40×10^{-3}
V^+Ex^+-specific modules	Type I interferon response	Nasal		262	1.20	2.93×10^{-2}	9.63×10^{-2}	1.34	2.94×10^{-3}	0.94	6.44×10^{-1}	1.01×10^{-8}
	Type I interferon response	Blood		51	1.28	4.03×10^{-3}	1.69×10^{-1}	1.32	4.90×10^{-3}	1.17	3.05×10^{-1}	3.70×10^{-2}
	Chemoattraction/cytotoxicity	Nasal	Macrophage	247	1.31	2.70×10^{-3}	1.18×10^{-2}	1.52	7.65×10^{-5}	0.94	6.82×10^{-1}	1.24×10^{-8}
	Heat shock proteins/stress response	Nasal		65	1.26	5.17×10^{-4}	3.36×10^{-3}	1.40	1.65×10^{-5}	0.99	9.49×10^{-1}	1.67×10^{-5}
	Cilia/IL-33 response	Nasal		409	1.53	6.51×10^{-5}	4.83×10^{-4}	1.82	1.79×10^{-6}	1.04	8.33×10^{-1}	2.22×10^{-4}
	Antigen-presenting cell co-stimulation	Nasal	Lymphocyte, macrophage	58	1.08	2.44×10^{-1}	4.09×10^{-1}	1.16	4.63×10^{-2}	0.91	3.91×10^{-1}	2.77×10^{-4}
	Antigen processing and presentation	Nasal	Macrophage	283	1.12	1.04×10^{-2}	4.17×10^{-2}	1.17	3.10×10^{-3}	1.02	7.89×10^{-1}	7.82×10^{-4}
	Unannotated	Nasal	Macrophage	65	1.12	5.97×10^{-2}	1.72×10^{-1}	1.17	2.65×10^{-2}	1.01	9.02×10^{-1}	3.66×10^{-3}
	Type 2 inflammation	Nasal	Eosinophil	242	1.21	2.73×10^{-3}	1.18×10^{-2}	1.26	2.40×10^{-3}	1.11	3.37×10^{-1}	1.05×10^{-2}
	Chromatin modification/regulation of gene expression	Nasal		144	0.93	1.67×10^{-1}	3.48×10^{-1}	0.88	4.34×10^{-2}	1.05	6.10×10^{-1}	2.32×10^{-3}
	Protein catabolism	Nasal		495	0.93	1.06×10^{-1}	2.50×10^{-1}	0.89	2.56×10^{-2}	1.03	6.99×10^{-1}	2.41×10^{-2}
V^-Ex^+-specific modules	Keratinization/epithelial development/cell–cell adhesion/tight junctions	Nasal	Epithelium	1,041	1.35	9.72×10^{-4}	5.62×10^{-3}	1.30	1.34×10^{-2}	1.45	1.65×10^{-2}	3.84×10^{-4}
	Tissue kallikreins/keratinization	Nasal		143	1.22	2.21×10^{-2}	8.22×10^{-2}	1.17	1.28×10^{-1}	1.34	5.00×10^{-2}	5.72×10^{-3}
	Squamous epithelium	Nasal	Epithelium	56	1.15	2.14×10^{-3}	1.11×10^{-2}	1.12	3.81×10^{-2}	1.23	1.01×10^{-2}	1.15×10^{-2}

In primary analysis, 13 (of 94 total) modules had significantly different expression when comparing the Ex^+ and Ex^- event groups in the full cohort (FDR < 0.05). Values correspond to the differences shown graphically in Fig. 3 and Supplementary Fig. 2. In secondary analysis, 20 modules had significantly different expression when comparing the V^+Ex^+, V^+Ex^-, V^-Ex^+, and V^-Ex^- subgroups (ANOVA, FDR < 0.05). These results are subdivided into virus-associated exacerbation modules that had different expression in the V^+Ex^+ and V^+Ex^- subgroups and nonviral exacerbation modules that had different expression in the V^-Ex^+ and V^-Ex^- subgroups. Values correspond to the differences shown graphically in Supplementary Fig. 5. Another five modules were significant in this comparison but only showed different expression between V^+ and V^- events with no difference according to exacerbation status; these modules are therefore excluded for brevity. The analysis included 247 unique nasal samples and 256 unique blood samples from 106 individuals who in total had 47 Ex^+ events and 107 Ex^- events. The event subgroups included 33 V^+Ex^+, 14 V^-Ex^+, 69 V^+Ex^-, and 38 V^-Ex^- events. Modules are listed according to their summary annotation from Supplementary Table 2, along with the source sample type (nasal or blood), cell type association based on deconvolution, and number of genes in the module. The first block of rows includes the six core modules that were significant in the primary and secondary analyses as well as the sensitivity analyses. The second block of rows includes those modules specific to V^+Ex^+ events. The third block of rows includes those modules specific to V^-Ex^+ events. Comparisons were performed by using a weighted linear model and empirical Bayes method, including terms for exacerbation status, cell percentages, presence or absence of virus, visit, and library sequencing depth with a random effect included for participant. For subgroup comparisons, presence or absence of virus was removed from the model and ANOVA was run to determine an overall significance value; pairwise comparisons were then run to determine significant differences across groups.

FDR < 0.05 across >99% of the iterations (Supplementary Fig. 4b,c). These results demonstrate the consistency of these six exacerbation-associated pathways within and across individuals; given the robust and reproducible nature of the involvement of these six modules in exacerbation, we refer to them as 'core' exacerbation modules.

Differential patterns of immune response in virus-associated and nonviral asthma exacerbations. The presence or absence of a respiratory virus was a primary source of variability in our transcriptome data. Consequently, we sought to define molecular signatures of exacerbation common or specific to virus-associated and nonviral events. Events were subgrouped as V^+Ex^+ ($n = 33$),

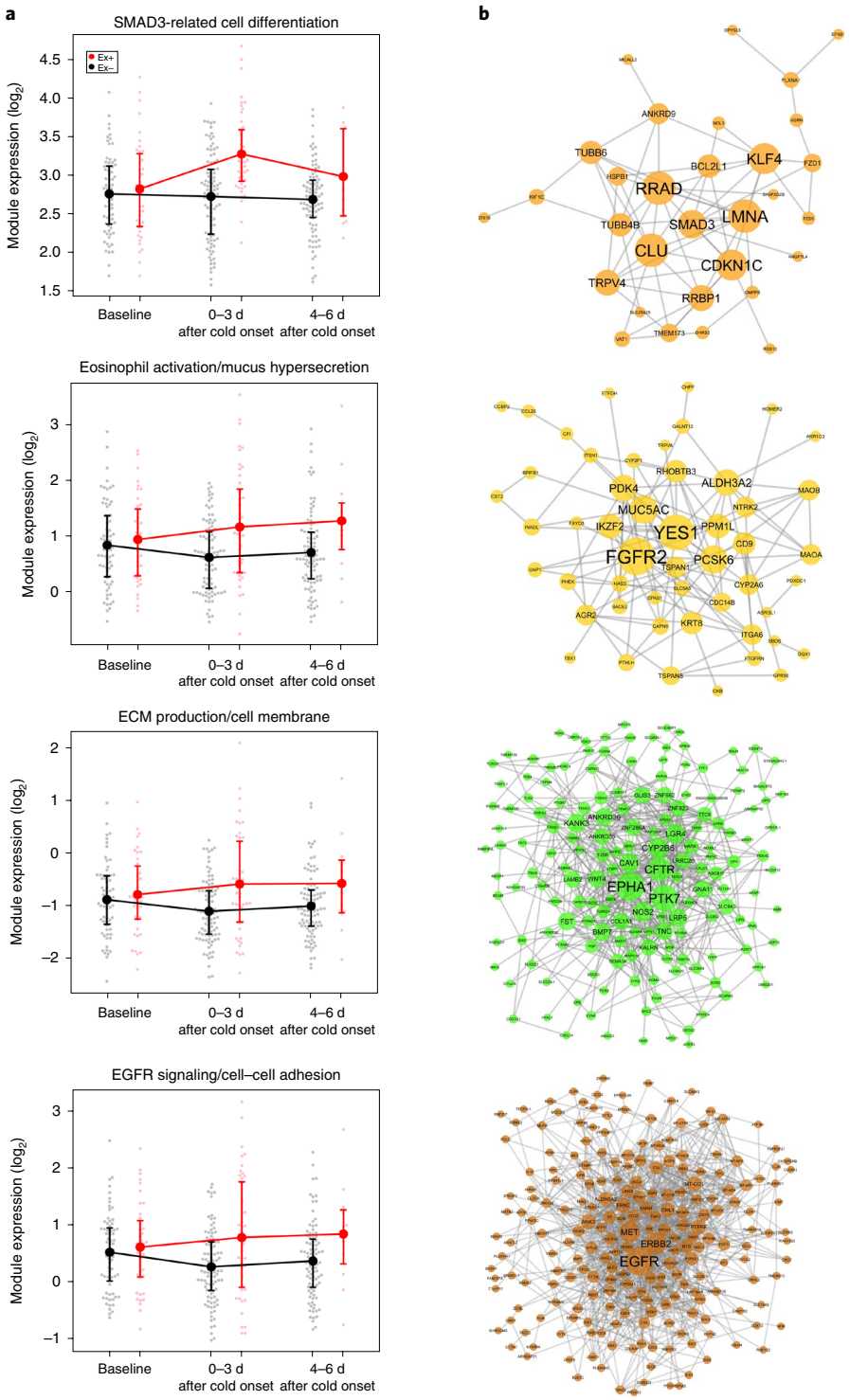


Fig. 3 | Four core exacerbation modules are upregulated in cold events that lead to exacerbations. a, Expression levels for the four most significantly different nasal gene expression modules that had increased expression in Ex^+ events as compared to Ex^- events. FDR values from top to bottom are 2.0×10^{-7}, 1.9×10^{-5}, 1.9×10^{-5}, and 5.6×10^{-5}. Expression levels represent the log base 2 of the geometric mean for normalized expression of all genes within the module. Shown are group mean values, interquartile range, and all data points. Sample sizes are as follows: Ex^+ at baseline, $n=38$; Ex^+ at 0–3 d, $n=44$; Ex^+ at 4–6 d, $n=11$; Ex^- at baseline, $n=68$; Ex^- at 0–3 d, $n=97$; Ex^- at 4–6 d, $n=95$. 18 participants who had one Ex^+ and one Ex^- event have the same measurement shown for Ex^+ at baseline and Ex^- at baseline. All measurements shown are otherwise biologically independent. Comparisons were performed by using a weighted linear model and empirical Bayes method, including terms for exacerbation status, cell percentages, presence or absence of virus, visit, and library sequencing depth with a random effect included for participant. **b**, Gene–gene associations for each of the four modules demonstrate significant interaction networks centered around key genes. Genes are represented as circular nodes, and known gene–gene interactions from STRING are shown as connecting edges. The size of each node is proportional to the number of interactions. The networks are drawn as force-directed graphs, meaning that genes toward the center have the greatest centrality within the network. STRING enrichment *P* values from top to bottom are 1.3×10^{-4}, 9.3×10^{-6}, 7.1×10^{-13}, and $<1.0\times10^{-16}$.

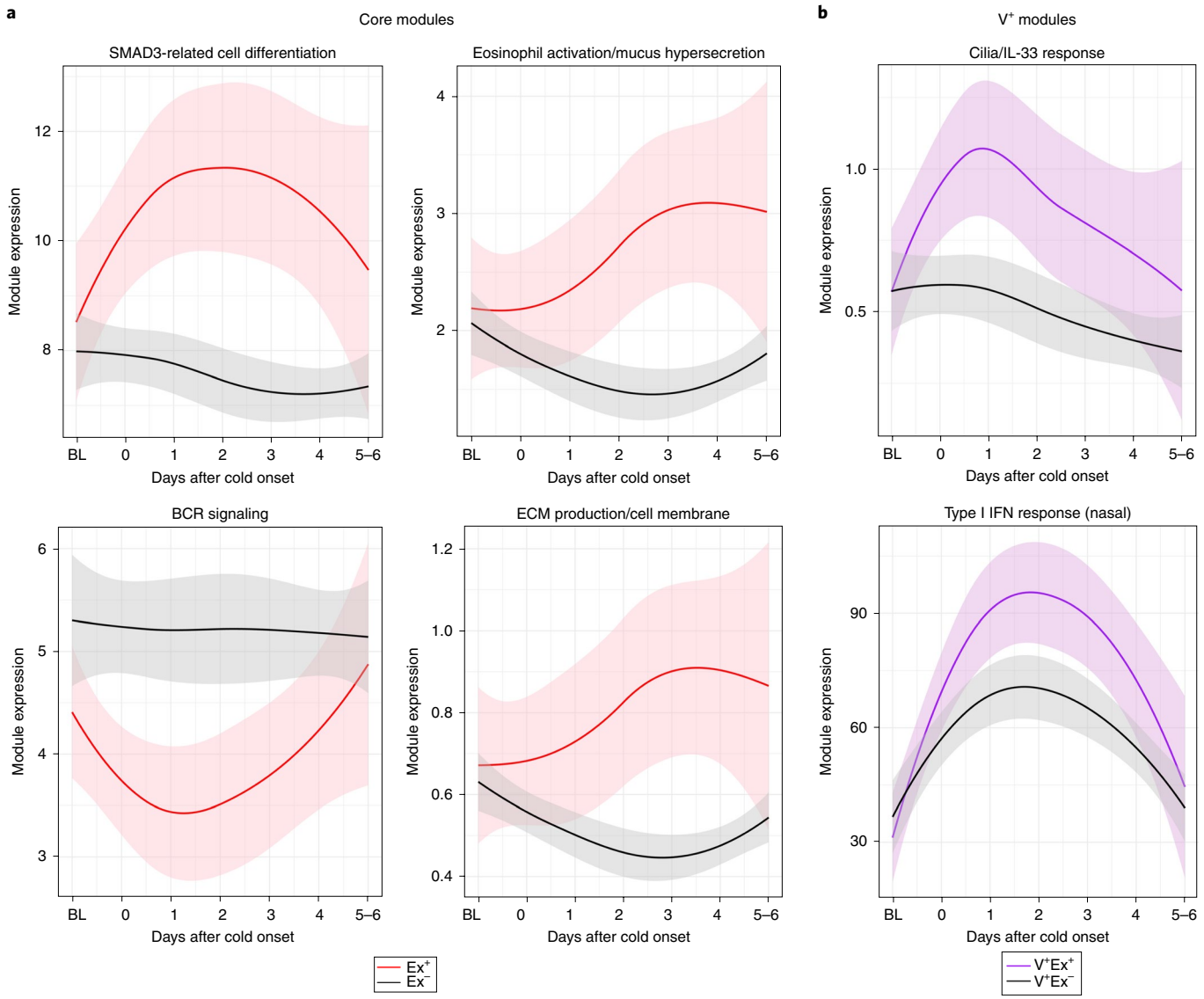


Fig. 4 | Longitudinal dynamics of module expression show patterns of sequential module activation. a, Among the core exacerbation modules, expression in the epithelial-associated 'SMAD3-related cell differentiation' module peaked early in the Ex^+ group (FDR = 4.7×10^{-5}) while expression in the 'BCR signaling' (FDR = 1.0×10^{-3}) and 'lymphocyte proliferation' (FDR = 4.7×10^{-5}; data not shown) modules nadired early. Expression in the 'eosinophil activation/mucus hypersecretion' (FDR = 1.3×10^{-3}), 'ECM production/cell membrane' (FDR = 2.1×10^{-3}), and 'EGFR signaling/cell–cell adhesion' (FDR = 4.0×10^{-3}; data not shown) modules increased over time in the Ex^+ group. Sample sizes are as follows: Ex^+, $n = 92$; Ex^-, $n = 244$; biologically independent samples. **b**, In V^+Ex^+ events, expression of genes in the 'cilia/IL-33 response' module is an early and specific event, with expression peaking in the first day and decreasing over time in V^+Ex^- events (FDR = 5.9×10^{-3}). The nasal 'type I IFN response' (FDR = 0.19) and 'HSPs/stress response' (FDR = 9.6×10^{-3}) modules and macrophage-associated 'chemoattraction/cytotoxicity' (FDR = 0.19) and 'APP' (FDR = 0.12) modules all had higher peak expression and area under the curve values in V^+Ex^+ events. The nasal 'type I IFN response' module is shown as a representative example for these modules, all of which exhibited similar dynamics. For nasal data, sample sizes were as follows: V^+Ex^+, $n = 67$; V^+Ex^-, $n = 157$; biologically in dependent samples. For blood data, sample sizes were as follows: V^+Ex^+, $n = 64$; V^+Ex^-, $n = 147$; biologically independent samples. Longitudinal data are plotted with a 95% confidence interval (shaded region) estimated by local second-degree polynomial regression. To assess significance, a linear model was fit with a B-spline basis for a polynomial spline with degree 2 for days after cold onset and ANOVA was run to determine whether exacerbation status was significant.

V^+Ex^- ($n = 69$), V^-Ex^+ ($n = 14$), and V^-Ex^- ($n = 38$). Similar virus types were observed in each V^+ subgroup; 60.6% of V^+Ex^+ and 50.7% of V^+Ex^- events were positive for either RV-A or RV-C (Supplementary Table 2).

After adjusting for cell percentages, 25 modules had significantly different expression among the groups (FDR < 0.05) (Table 2). Expression of the six previously defined core exacerbation modules was significantly different in both the V^+Ex^+ and V^-Ex^+ subgroups in comparison to the respective Ex^- subgroups and the expression magnitudes were statistically similar in the Ex^+ subgroups, further confirming that these pathways are core to exacerbations, regardless of the apparent subgroup or inciting event (Supplementary Fig. 5a).

Another seven nasal modules and one blood module had expression specifically elevated in V^+Ex^+ events as compared to the other three subgroups (FDR < 0.05), demonstrating specificity of these pathways to virus-triggered exacerbations. The V^+Ex^+-specific modules represent numerous inflammatory pathways not seen in V^- events. These included a 'type I interferon (IFN) response' module

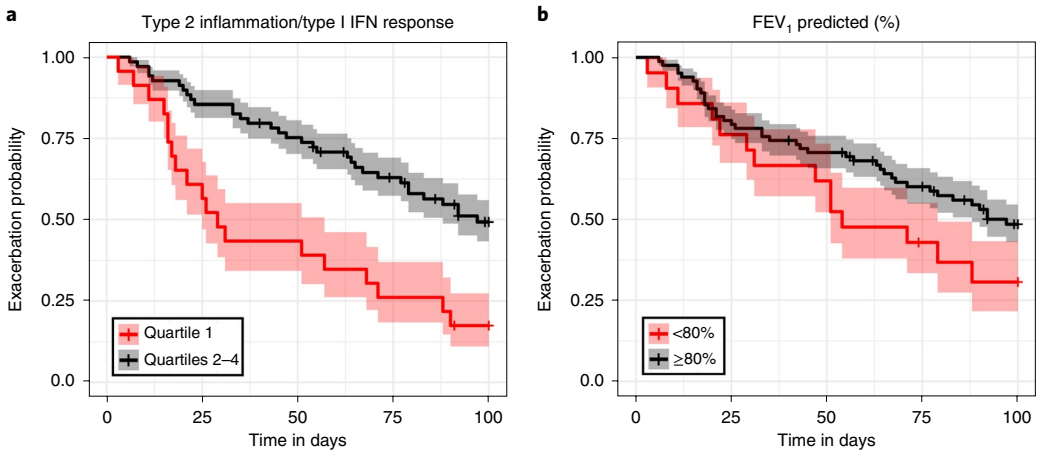


Fig. 5 | The ratio of expression for the 'type 2 inflammation' and 'type I IFN response' modules predicts exacerbation risk. a, Kaplan–Meier analysis demonstrates that the ratio of expression in the nasal 'type 2 inflammation' module to expression in the nasal 'type I IFN response' module at the baseline healthy visit is predictive of the time to a participant's next asthma exacerbation. Individuals in the highest quartile for this expression ratio (quartile 1; $n=23$) had a significantly shorter time to exacerbation than individuals in the lower three quartiles (quartiles 2–4; $n=69$) ($P=2.6\times10^{-4}$; the lower three quartiles were not statistically different from one another), and they had >50% risk of an exacerbation in the first 30 d after the baseline visit. **b**, In comparison, Kaplan–Meier analysis demonstrates that baseline FEV_1 predicted (%) does not significantly predict time to the next exacerbation, when individuals are split into those with FEV_1 predicted <80% ($n=21$) and those with FEV_1 predicted ≥80% ($n=82$) ($P=0.11$). The same is true when individuals are split into quartiles according to FEV_1 predicted ($P=0.3$; data not shown). All samples shown are biologically independent. Shaded regions represent 1 s.d. log confidence intervals. Plus symbols represent censored observations.

in both nasal and blood samples, which contained *IRF7*, *STAT1*, *STAT2*, and numerous antiviral effector molecules; a respiratory epithelium–associated 'cilia/IL-33 response' pathway including *IL33*, *FOXJ1*, *TLR3*, and *CDHR3*, in association with genes of ciliogenesis, cell–cell adhesion, and ECM adhesion; and a nasal 'heat shock proteins (HSPs)/stress response' module. They also included the nasal macrophage-associated modules 'chemoattraction/cytotoxicity', 'antigen processing and presentation (APP)', and 'antigen-presenting cell (APC) co-stimulation'. Each of these modules had the highest expression in V^+Ex^+ events in comparison to the three other groups (FDR < 0.05), with expression in some modules modestly elevated in V^+Ex^- as compared to V^- events (both V^-Ex^+ and V^-Ex^-) (Supplementary Fig. 5b). These findings demonstrate a multifaceted immune response to viruses that is more intense in viral infections that lead to an exacerbation. Two nasal modules had expression that was specifically decreased in V^+Ex^+ events in comparison to the other three groups; these modules were annotated as 'protein catabolism' and 'chromatin modification/regulation of gene expression'. One other module had significantly different expression between V^+Ex^+ and V^+Ex^- events, 'type 2 inflammation', which includes the cytokines *IL4*, *IL5*, and *IL13*, the receptors *IL5RA* and *IL3RA*, and the transcription factors *GATA1* and *GATA2*. This module had the lowest expression in V^+Ex^- events, in comparison to V^+Ex^+ events (two-group FDR < 0.05) or the V^- subgroups, but expression was not statistically different between the V^+Ex^+ and V^- groups, suggesting that a relative lack of type 2 inflammation in the airway corresponds to resolution of a viral cold without development of an asthma exacerbation.

Another three nasal modules had expression specifically elevated only in the V^-Ex^+ subgroup in comparison to the other three groups (FDR < 0.05), demonstrating specificity of these pathways to nonviral exacerbations. These were each highly associated with squamous percentages. These modules included the largest module in the repertoire annotated as 'keratinization/epithelial development/cell–cell adhesion/tight junctions', which is composed of 1,041 genes, is enriched for numerous barrier functions, and includes *CDH1* (E-cadherin) and *FLG* (filaggrin). The other two modules were 'tissue kallikreins/keratinization', which includes seven tissue kallikreins and several proteins and enzymes responsible for cornification, and 'squamous epithelium' (Supplementary Fig. 5c). These findings point to distinct and partially squamous cell–specific initiation of asthma exacerbations triggered in the absence of virus.

These results confirm that the set of core modules are common to both virus-associated and nonviral exacerbations while demonstrating that there is a distinct set of respiratory epithelium and inflammatory cell pathways involved in virus-associated exacerbations in contrast to the predominantly squamous cell pathways underlying nonviral exacerbations.

Sequential immune activation underlying the development of an asthma exacerbation. We next investigated the kinetics of module expression during illnesses in the Ex^+ and Ex^- groups relative to baseline values by using local regression to incorporate the timing of observed transcriptional changes relative to the onset of cold symptoms. Expression of genes in the 'SMAD3-related cell differentiation' module was significantly increased in the Ex^+ group within 1 d of cold onset and was sustained for several days throughout the event (FDR < 0.001; Fig. 4a). Similarly, expression in the 'lymphocyte proliferation' and 'BCR signaling' modules nadired rapidly in the Ex^+ group, within 1–2 d of cold symptom onset (FDR < 0.001). These early changes in immune response were followed by slower elevation and later peaks (at approximately day 4) in the expression of genes in the 'eosinophil activation/mucus hypersecretion', 'ECM production/cell membrane', and 'EGFR signaling/cell–cell adhesion' modules, all of which showed a steady increase in expression in the Ex^+ group in comparison to baseline while showing decreased expression in the Ex^- group (FDR < 0.05; Fig. 4a). These later-peaking modules appear to represent 'effector' functions related to the pathophysiology of exacerbations, and these dynamics demonstrate the sequential pathway activation underlying exacerbation events.

In the V^+Ex^+ subgroup, we observed that expression in the 'cilia/IL-33 response' module peaked rapidly, within 1 d of cold onset, whereas expression in the V^+Ex^- subgroup of this module decreased throughout the event (FDR < 0.05; Fig. 4b). Expression

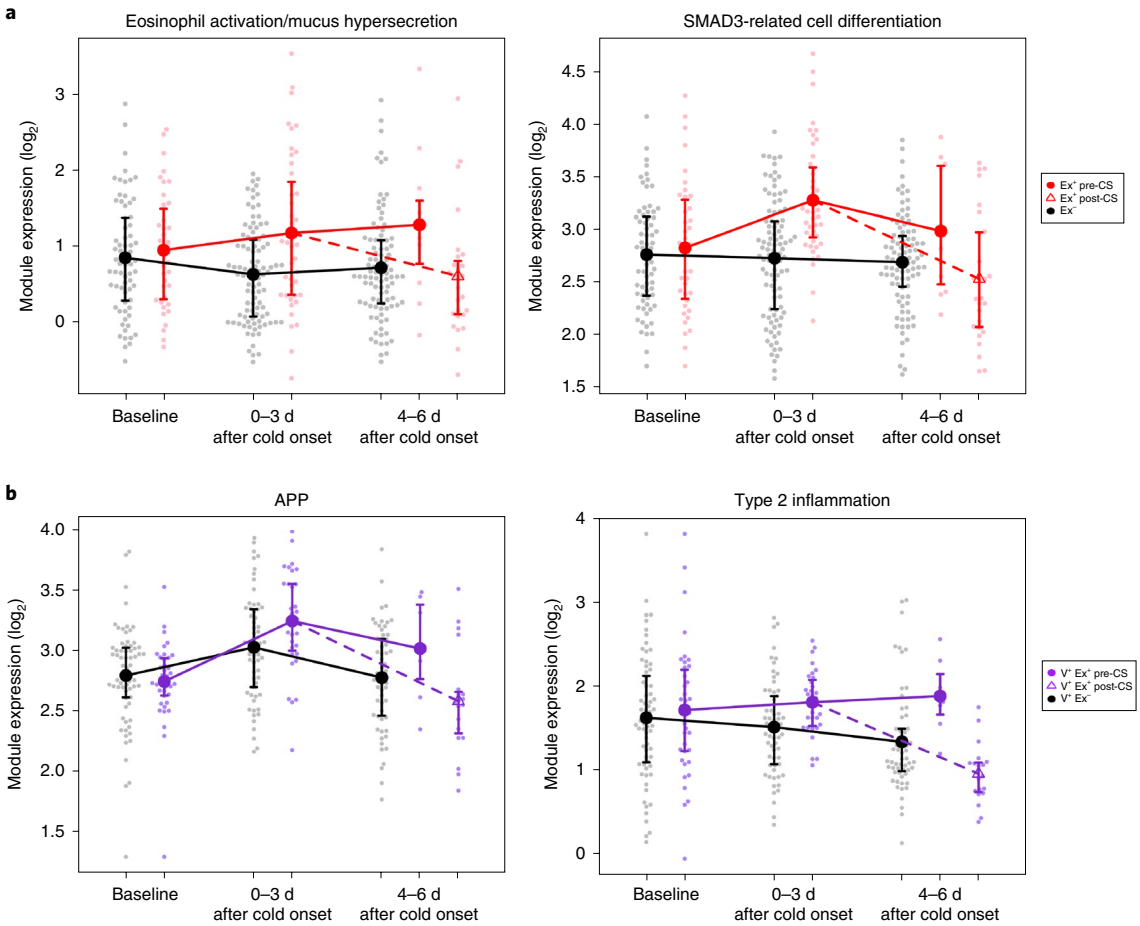


Fig. 6 | Systemic corticosteroids affect only a subset of exacerbation modules. a, Among the core exacerbation modules, 'eosinophil activation/mucus hypersecretion' and 'SMAD3-related cell differentiation' showed decreased expression after systemic corticosteroids were started (post-CS) in comparison to expression levels before systemic corticosteroids were started (pre-CS) (FDR = 4.6×10^{-2} and FDR = 7.1×10^{-2}, respectively). Sample sizes were as follows: Ex^+ at baseline, $n = 38$; Ex^+ at 0–3 d, $n = 44$; Ex^+ pre-CS at 4–6 d, $n = 11$; Ex^+ post-CS at 4–6 d, $n = 27$; Ex^- at baseline, $n = 68$; Ex^- at 0–3 d, $n = 97$; Ex^- at 4–6 d, $n = 95$. **b**, Among the V^+Ex^+ exacerbation modules, 'APP' (FDR = 5.0×10^{-2}) and 'type 2 inflammation' (FDR = 5.1×10^{-3}) showed decreased expression after systemic corticosteroids were started. Sample sizes were as follows: V^+Ex^+ at baseline, $n = 38$; V^+Ex^+ at 0–3 d, $n = 31$; V^+Ex^+ pre-CS at 4–6 d, $n = 7$; V^+Ex^+ post-CS at 4–6 d, $n = 19$; V^+Ex^- at baseline, $n = 68$; V^+Ex^- at 0–3 d, $n = 61$; V^+Ex^- at 4–6 d, $n = 59$. 18 participants who had one Ex^+ and one Ex^- event have the same measurement shown for Ex^+ at baseline and Ex^- at baseline. All measurements shown are otherwise biologically independent. Comparisons were performed by using a weighted linear model and empirical Bayes method, including terms for status (Ex^-, Ex^+ pre-CS, and Ex^+ post-CS), cell percentages, presence or absence of virus, visit, and library sequencing depth with a random effect included for participant. Contrasts were tested for status within and across visits. The term for presence or absence of virus was removed in the analysis limited to V^+ events.

of the 'type 2 inflammation' module stayed at a stable level in the V^+Ex^+ subgroup, whereas it steadily decreased in the V^+Ex^- subgroup (FDR < 0.05). The nasal 'type I IFN response' and 'HSPs/stress response' modules and the macrophage-associated 'chemoattraction/cytotoxicity' and 'APP' pathways each showed peak expression between days 1–3 and expression was significantly increased during both V^+Ex^+ and V^+Ex^- events; however, V^+Ex^+ events had higher peak expression levels and higher areas under the curve, in line with sustained expression of the genes in these inflammatory pathways at increased levels in V^+Ex^+ as compared to V^+Ex^- events (FDR values of 0.01–0.19; Fig. 4b). In the V^-Ex^+ subgroup, we observed that the three squamous-associated pathways had elevated expression on days 1–3 and expression then returned toward baseline levels; however, the model for this comparison was not statistically significant (FDR > 0.25).

Through this longitudinal analysis, we discerned kinetic models of the molecular events underpinning cold events that resolves versus those that lead to an exacerbation.

Baseline immune state represents a risk factor for exacerbation. We next investigated how levels of module expression at the baseline healthy visit influence subsequent risk of an asthma exacerbation. We compared module expression levels to the time from baseline to the first asthma exacerbation. Four nasal modules were significantly associated with time to exacerbation ($P < 0.05$; Supplementary Table 3). To determine the optimal model, we performed forward stepwise regression with the significant modules. Two of these modules, 'type 2 inflammation' and 'type I IFN response', each independently contributed significantly to overall model fit (beta coefficient, $P < 0.05$), and the model incorporating both of these modules showed significantly better fit (ANOVA, $P < 0.05$) than either univariate model. Specifically, higher expression of the 'type 2 inflammation' module and lower expression of the 'type I IFN response' module were associated with a shorter time to exacerbation. When related to exacerbation probability, an elevated ratio of 'type 2 inflammation' to 'type I IFN response' expression corresponded to a highly significant risk for

exacerbation in the short term, with individuals in the highest quartile having a significantly shorter time to exacerbation than individuals in the lower three quartiles ($P < 0.001$) and >50% risk of an exacerbation in the first 30 d after the baseline visit (Fig. 5a). This expression ratio was a better predictor of short-term exacerbation risk than any of the collected clinical variables, such as baseline FEV_1 predicted (%) (Fig. 5b). These results show that a baseline gene expression pattern constituting high expression of nasal type 2 inflammation genes and low expression of type I IFN response genes is a robust predictor of short-term exacerbation risk.

Effects of systemic corticosteroid treatment on the immune response. To examine the effects of systemic corticosteroid administration on immune responses, we studied 30 samples that had been collected after systemic corticosteroid intervention in Ex^+ events in the period 4–6 d (average 5.2 d) after the onset of cold symptoms in comparison to samples collected in the same time window but prior to systemic corticosteroid intervention (average 4.8 d). Corticosteroids induced broad changes in immune pathways, altering the expression of 11 nasal modules and nasal eosinophil and neutrophil percentages, as well as the expression of 6 blood modules and blood eosinophil and neutrophil percentages (FDR < 0.05; Supplementary Table 4). Notably, expression of only a subset of the core exacerbation-associated modules was significantly altered by systemic corticosteroid use. Expression of genes in the 'eosinophil activation/mucus hypersecretion' and 'SMAD3-related cell differentiation' modules returned to levels equivalent to baseline after initiation of systemic corticosteroids (Fig. 6a); however, the other four core exacerbation modules did not show a significant change in expression with systemic corticosteroids. Among the V^+Ex^+-specific modules, systemic corticosteroid use led to significant reduction of expression only in the 'APP' and 'type 2 inflammation' modules (Fig. 6b). These results show that systemic corticosteroid treatment affects only a subset of the pathways involved in asthma exacerbations while also affecting multiple other biological processes.

A modular network graph demonstrates the patterns underlying exacerbation types. To facilitate interpretation of results and develop a platform for future comparisons and studies, we generated a cell–module bipartite network by using the entire dataset (Supplementary Fig. 6a). This network visually clusters co-associated modules and cell types to help identify co-occurring molecular events, leveraging the high degree of clustering among modules and cell differentials (Supplementary Fig. 6b).

By using this framework, the similarities and differences between V^+Ex^+ and V^-Ex^+ events are evident (Fig. 7a,b). The upregulated core exacerbation modules can be seen as common to both Ex^+ groups, clustering in close association with airway epithelial cells and eosinophils as expected, but also with peripheral blood eosinophils. Specific to V^+Ex^+ events, the upregulated 'cilia/IL-33 response' and 'type 2 inflammation' modules cluster together, demonstrating their close co-association, and near the upregulated core modules. Furthermore, we observe a distinct block of modules representing the type I IFN response as well as APC-related responses that are specific to V^+Ex^+ events, closely associating with nasal macrophages and lymphocytes as well as blood monocytes. In contrast, in V^-Ex^+ events, we observe that the same core modules were upregulated but cluster in association with the three upregulated squamous-associated modules.

The same network can be used to investigate other signals within the dataset; systemic corticosteroid treatment led to a decrease in eosinophils and downregulation of eosinophil-associated modules and two core exacerbation modules, without impacting the other exacerbation-associated modules (Fig. 7c). Also visible is the increase in airway neutrophils and expression of neutrophil-associated modules.

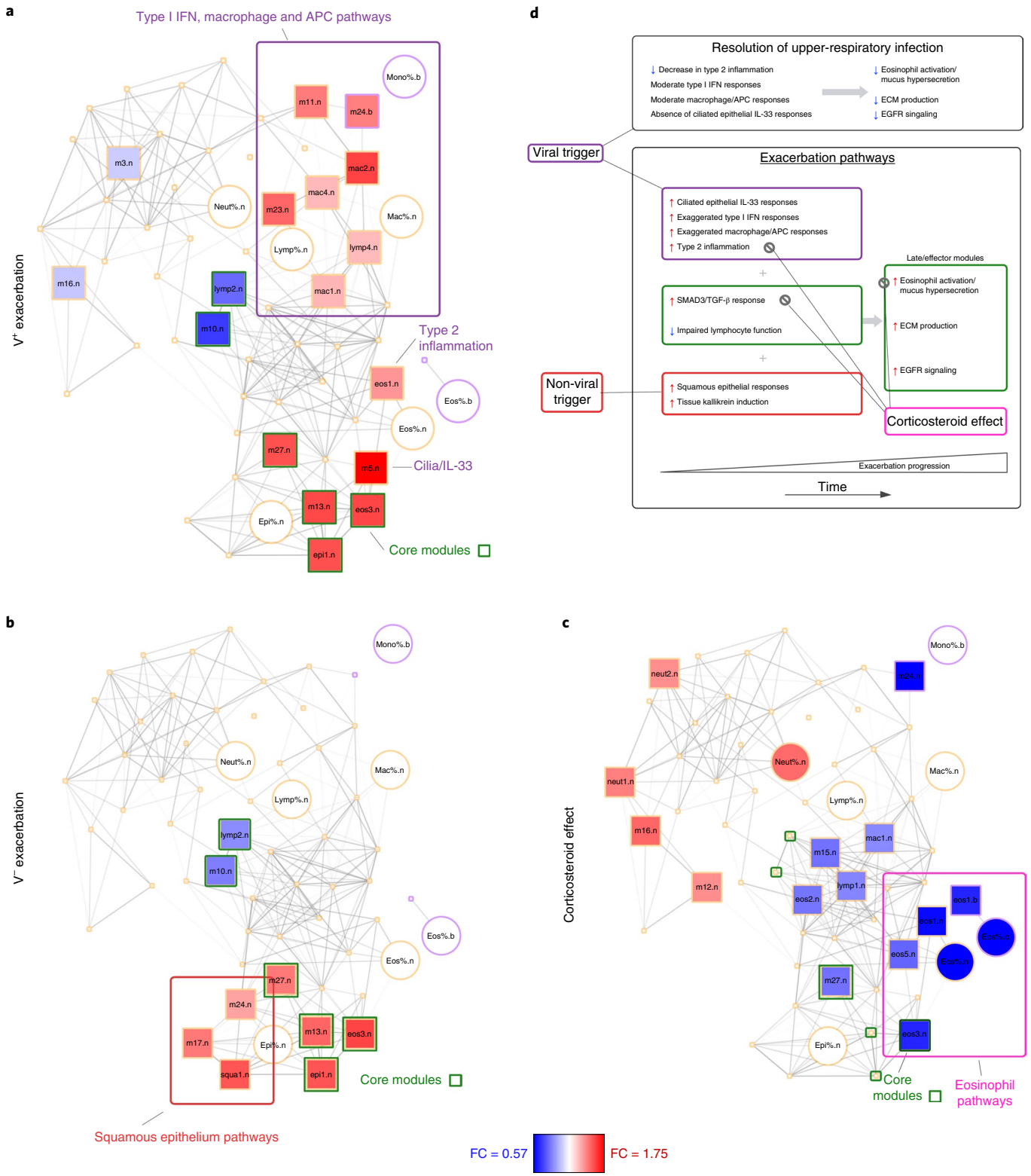
In conjunction with the previous results, this network analysis facilitates development of a model of the molecular events underpinning a cold event that resolves versus one that leads to an exacerbation and the effects of therapeutic intervention with systemic corticosteroids (Fig. 7d).

Discussion

Determining pathways by which asthma exacerbations occur in relation to distinct triggers and defining why some upper-respiratory illnesses provoke an exacerbation whereas others do not are critical areas of interest in asthma. We explore these areas through longitudinal transcriptome network analyses of nasal and blood samples from a cohort of exacerbation-prone children. Previous asthma transcriptomic studies have been more limited, defining gene expression levels in relation to stable or active disease sta tes[11–13,25–30]. Our results provide systems-level, data-driven determination of a complex network of molecular responses underpinning asthma exacerbations in comparison to both cold symptoms without exacerbation and stable disease. They provide important granularity for multiple distinct biological pathways, the relationships among these pathways, and their relative kinetics leading to an exacerbation.

Our findings are consistent with several established mechanisms of asthma exacerbation, confirm their importance in a highly relevant human study, and add appreciable new information to advance insight into exacerbation pathogenesis. An important result is the

Fig. 7 | Network overview of modular expression patterns demonstrates co-associated biological pathways. a, The signal for V^+Ex^+ events includes a cluster of the core exacerbation modules (outlined in green squares), as well as a cluster of the type I IFN, macrophage-associated, and APC-related modules (dark purple box), the 'cilia/IL-33 response' module, and the 'type 2 inflammation' module (dark purple labels). **b**, The signal for V^-Ex^+ events also includes a cluster of the core exacerbation modules as well as a cluster of the squamous-associated modules (dark orange box). **c**, The signal from corticosteroid treatment affects multiple areas of the network, most notably resulting in a marked decrease in clustered nasal and blood eosinophil cell percentages and module expression (magenta box). The bipartite network in each panel demonstrates the co-associations among nasal (light orange) and blood (light purple) modules (squares) and cell percentages (circles). Edges represent significant positive Pearson correlations >0.5, and darker edges indicate higher correlations. Nodes are clustered according to their interconnectedness. For each comparison, significant differences in module expression and cell percentages are colored red (increased) or blue (decreased), with color intensity corresponding to fold change. Modules whose expression was not significantly different (FDR ≥ 0.07) are shown as points rather than squares to simplify the diagram. **d**, The diagram summarizes the main results of our analysis, showing the key pathways observed related to viral and nonviral exacerbation events. Shown in two green boxes are the annotated molecular functions of the core exacerbation modules, related to upregulation of respiratory epithelium, squamous, and eosinophil responses and downregulation of lymphocyte responses, common to both viral and nonviral cold events that progress to an asthma exacerbation. These are separated into early-response pathways (left) and late/effector pathways (right) on the basis of the results of the longitudinal analysis. Shown in the dark purple box is the distinct set of respiratory epithelium– and inflammatory cell–associated pathways observed only in virus-associated exacerbations. Shown in the dark orange box are the predominantly squamous cell–associated pathways observed only in nonviral exacerbations. The pathways affected by systemic corticosteroid use are labeled. The contrasting responses observed in resolution of a viral upper-respiratory infection are shown at the top.



identification of the relevance of epithelial-associated SMAD3 signaling as a key early event in exacerbations. SMADs are signaling proteins that are a surrogate for TGF-β activity[31] and are known to have a central role in airway inflammation and remodeling in asthma; SMAD3 is required for airway remodeling in mice[32]. We observe that this pathway is upregulated in exacerbations regardless of the apparent trigger and upregulation is an early, central event in exacerbation development. This observation may represent a mechanism central to the link between repeated exacerbations and decline in lung function[33], with TGF-β activity inducing SMAD3 signaling within the epithelial–mesenchymal-trophic unit to enhance airway remodeling[34].

Our results support the importance of IL-33 and type 2 inflammation in asthma exacerbations[8]. IL-33 is an alarmin synthesized in response to epithelial damage, including from respiratory viruses[35], and amplifies innate immune pathways including type 2 inflammation[8,36]. Although pathogenic in asthma, IL-33 also has a key role in tissue repair[35], functions we see linked within the 'cilia/

IL-33 response' module, which contains *IL33*, *FOXJ1*, and genes related to ciliogenesis, cell–cell adhesion, and ECM adhesion. This network also contains *TLR3*, which encodes a viral sensor that induces IL-33 expression[37], and *CDHR3*, a key susceptibility locus for childhood asthma[38] that encodes the protein by which RV-C enters the epithelium, thus suggesting that the IL-33-mediated protection/repair process may leave the epithelium more susceptible to repeated infections. Upregulation of this module was closely associated with expression of the type 2 inflammation module, recapitulating the known interplay between these two key asthma mechanisms[8].

We observe asthma exacerbations that occur in the absence of a virus and are characterized by induction of keratinization, epithelial barrier pathways, and tissue kallikreins, which may represent metaplasia of the epithelium toward keratinized squamous cells, as observed with noxious exposures such as cigarette smoke[39]. Kallikreins generate kinins, including bradykinin, a key mediator of multiple acute responses in asthma that also impacts airway remodeling[40,41]. Notably, our findings directly relate features of airway epithelial barrier dysfunction to epithelial kallikrein induction in asthma exacerbations. Although we could not identify a definite etiology for these exacerbations, smoke, pollutants, or allergens are possible causes, as each of these can disrupt epithelial integrity and in animal models can lead to production of bradykinin[42,43].

We demonstrate that high type 2 inflammation and low type I IFN response gene expression in nasal samples at baseline predicts short-term exacerbation risk, building on previous results showing that nasal *IL13* expression is associated with exacerbation likelihood[30]. This suggests a specific 'at-risk' immune state, when a child is highly susceptible to exacerbation. Furthermore, our results provide insight into conflicting data regarding antiviral IFN responses in asthma from earlier studies, in which some have shown relative deficiency in IFN responses[44,45] whereas others have shown normal or excessive responses[46,47]. Our data show that both observations may be valid; IFN responses may not be consistently diminished or elevated, but rather dysregulated. Low IFN signaling at baseline may enable viral replication and/or epithelial damage during the early phase of infection, which secondarily induces an exaggerated IFN response promoting exacerbation. We show that systemic corticosteroid treatment affects only a subset of exacerbation-associated pathways, in particular those related to eosinophil activation and mucus hypersecretion, type 2 inflammation, and epithelial SMAD3 signaling. The failure of corticosteroids to moderate other pathways, including EGFR signaling and ECM production, may help to explain why corticosteroid treatment has limited efficacy in preventing lung function decline[48]. This result also provides an example of definition of the in vivo molecular effects of therapeutics, which is critical to improving interventional approaches.

A major strength of our study is its prospective design. We captured cold symptoms and natural viral infections as opposed to using viral inoculation or an in vitro model. By using daily symptom monitoring, we captured events at the beginning of symptoms prior to both clinical exacerbation and initiation of systemic corticosteroids, while also capturing follow-up samples to investigate progression of the immune response and the effect of systemic corticosteroids. By using transcriptome network analysis, we were able to investigate our aims in an unbiased manner, leading to the discovery of new findings while also confirming and expanding upon anticipated results. Through serially combining two powerful network analysis techniques, cell deconvolution and WGCNA, in a novel algorithm, we distilled a massive dataset into coherent, statistically rigorous, and biologically meaningful conclusions.

A limitation to our approach was the use of upper-airway samples as a proxy for lower-airway events. Previous work has demonstrated that nasal epithelial expression profiles are globally similar to those of the lower airway[30] and are specifically similar in key pathways, including aspects of remodeling[49], but there is also evidence of differences in certain immune pathways, including interferon responses[50]. It is possible that we have missed aspects of the lower-airway inflammatory response by not directly sampling the lower airway during illness. However, the clear differences between Ex^+ and Ex^- events and the corresponding decrease in pulmonary functions suggest that our findings reflect molecular pathways directly relevant to the pathophysiology of exacerbations. Paired upper- and lower-airway samples collected in future work can build upon our results. By collecting lavage from the upper airway, we obtained samples that were mixtures of inflammatory and epithelial cells. Given the clear importance of respiratory epithelium in viral infections and asthma, future studies could use epithelial brushings as a means to potentially further define the role of the epithelium in exacerbations.

Our methodology enabled us to obtain significant data and meaningful results from a sample readily collected in a clinical setting with minimal invasiveness. This approach was successfully and safely implemented in children and adolescents from a high-risk population, in which lower-airway sampling is not feasible during an emerging asthma exacerbation. Furthermore, our modular repertoire and network visualization provide a platform to test new comparisons and analyze new datasets. The current study did not investigate specific interventions and is not intended to definitively determine causation. This in vivo platform can be used to investigate the granularity of specific exacerbation subtypes and the effects of therapies anticipated to block specific components of exacerbation pathways.

In conclusion, we have demonstrated the application of transcriptome network analyses as a robust methodology to define networks of immune responses underpinning asthma exacerbations in children. This strategy revealed coordinated epithelial and eosinophil pathways that are upregulated as well as lymphocyte pathways that are downregulated in both virus-associated and nonviral exacerbations. We also found distinct responses in viral and nonviral exacerbation events and identified baseline characteristics associated with susceptibility to exacerbations. Finally, we have developed a new sample collection and analysis platform that can be readily integrated into translational asthma research. Collectively, we anticipate that these approaches will advance understanding of both the efficacy of current and novel intervention strategies and unresponsiveness to these strategies in some treated individuals, thereby leading to a clearer and more comprehensive understanding of asthma exacerbations and how to prevent and treat them.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41590-019-0347-8.

Received: 10 July 2018; Accepted: 11 February 2019;
Published online: 08 April 2019

References

1. Johnston, S. L. et al. Community study of role of viral infections in exacerbations of asthma in 9–11 year old children. *BMJ* **310**, 1225–1229 (1995).
2. Busse, W. W., Lemanske, R. F. Jr & Gern, J. E. Role of viral respiratory infections in asthma and asthma exacerbations. *Lancet* **376**, 826–834 (2010).
3. Olenec, J. P. et al. Weekly monitoring of children with asthma for infections and illness during common cold seasons. *J. Allergy Clin. Immunol.* **125**, 1001–1006 (2010).
4. Busse, W. W. et al. Randomized trial of omalizumab (anti-IgE) for asthma in inner-city children. *N. Engl. J. Med.* **364**, 1005–1015 (2011).
5. Pavord, I. D. et al. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. *Lancet* **380**, 651–659 (2012).

6. Castro, M. et al. Dupilumab efficacy and safety in moderate-to-severe uncontrolled asthma. *N. Engl. J. Med.* **378**, 2486–2496 (2018).
7. Drazen, J. M. & Harrington, D. New biologics for asthma. *N. Engl. J. Med.* **378**, 2533–2534 (2018).
8. Jackson, D. J. et al. IL-33-dependent type 2 inflammation during rhinovirus-induced asthma exacerbations in vivo. *Am. J. Respir. Crit. Care Med.* **190**, 1373–1382 (2014).
9. Farne, H. A. & Johnston, S. L. Immune mechanisms of respiratory viral infections in asthma. *Curr. Opin. Immunol.* **48**, 31–37 (2017).
10. Chaussabel, D. & Baldwin, N. Democratizing systems immunology with modular transcriptional repertoire analyses. *Nat. Rev. Immunol.* **14**, 271–280 (2014).
11. Woodruff, P. G. et al. Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc. Natl Acad. Sci. USA* **104**, 15858–15863 (2007).
12. Bosco, A., Ehteshami, S., Panyala, S. & Martinez, F. D. Interferon regulatory factor 7 is a major hub connecting interferon-mediated responses in virus-induced asthma exacerbations in vivo. *J. Allergy Clin. Immunol.* **129**, 88–94 (2012).
13. Wesolowska-Andersen, A. et al. Dual RNA-seq reveals viral infections in asthmatic children without respiratory illness which are associated with changes in the airway transcriptome. *Genome Biol.* **18**, 12 (2017).
14. Berube, J. C. & Bosse, Y. Future clinical implications emerging from recent genome-wide expression studies in asthma. *Expert Rev. Clin. Immunol.* **10**, 985–1004 (2014).
15. Bonser, L. R. & Erle, D. J. Airway mucus and asthma: the role of MUC5AC and MUC5B. *J. Clin. Med.* **6**, E112 (2017).
16. Paul, M. K. et al. Dynamic changes in intracellular ROS levels regulate airway basal stem cell homeostasis through Nrf2-dependent Notch signaling. *Cell Stem Cell* **15**, 199–214 (2014).
17. Kim, J. T., Gleich, G. J. & Kita, H. Roles of CD9 molecules in survival and activation of human eosinophils. *J. Immunol.* **159**, 926–933 (1997).
18. Gudbjartsson, D. F. et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* **41**, 342–347 (2009).
19. Loffredo, L. F. et al. Beyond epithelial-to-mesenchymal transition: common suppression of differentiation programs underlies epithelial barrier dysfunction in mild, moderate, and severe asthma. *Allergy* **72**, 1988–2004 (2017).
20. Reeves, S. R. et al. Asthmatic airway epithelial cells differentially regulate fibroblast expression of extracellular matrix components. *J. Allergy Clin. Immunol.* **134**, 63–670 (2014).
21. Hamid, Q. et al. Induction of nitric oxide synthase in asthma. *Lancet* **342**, 1510–1513 (1993).
22. Coulthard, M. G. et al. Eph/Ephrin signaling in injury and inflammation. *Am. J. Pathol.* **181**, 1493–1503 (2012).
23. Amishima, M. et al. Expression of epidermal growth factor and epidermal growth factor receptor immunoreactivity in the asthmatic human airway. *Am. J. Respir. Crit. Care Med.* **157**, 1907–1912 (1998).
24. Takeyama, K., Fahy, J. V. & Nadel, J. A. Relationship of epidermal growth factor receptors to goblet cell production in human bronchi. *Am. J. Respir. Crit. Care Med.* **163**, 511–516 (2001).
25. Bigler, J. et al. A severe asthma disease signature from gene expression profiling of peripheral blood from U-BIOPRED cohorts. *Am. J. Respir. Crit. Care Med.* **195**, 1311–1320 (2017).
26. Kuo, C.-H. S. et al. A transcriptome-driven analysis of epithelial brushings and bronchial biopsies to define asthma phenotypes in U-BIOPRED. *Am. J. Respir. Crit. Care Med.* **195**, 443–455 (2017).
27. Persson, H. et al. Transcriptome analysis of controlled and therapy-resistant childhood asthma reveals distinct gene expression profiles. *J. Allergy Clin. Immunol.* **136**, 638–648 (2015).
28. Sordillo, J. E. et al. Genome-wide expression profiles identify potential targets for gene–environment interactions in asthma severity. *J. Allergy Clin. Immunol.* **136**, 885–892 (2015).
29. Subrata, L. S. et al. Interactions between innate antiviral and atopic immunoinflammatory pathways precipitate and sustain asthma exacerbations in children. *J. Immunol.* **183**, 2793–2800 (2009).
30. Poole, A. et al. Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. *J. Allergy Clin. Immunol.* **133**, 670–678 (2014).
31. Al-Alawi, M., Hassan, T. & Chotirmall, S. H. Transforming growth factor β and severe asthma: a perfect storm. *Respir. Med.* **108**, 1409–1423 (2014).
32. Le, A. V. et al. Inhibition of allergen-induced airway remodeling in Smad 3-deficient mice. *J. Immunol.* **178**, 7310–7316 (2007).
33. O'Byrne, P. M., Pedersen, S., Lamm, C. J., Tan, W. C. & Busse, W. W. Severe exacerbations and decline in lung function in asthma. *Am. J. Respir. Crit. Care Med.* **179**, 19–24 (2009).
34. Holgate, S. T. et al. Asthma. *Nat. Rev. Dis. Primers* **1**, 15025 (2015).
35. Liew, F. Y., Girard, J. P. & Turnquist, H. R. Interleukin-33 in health and disease. *Nat. Rev. Immunol.* **16**, 676–689 (2016).
36. Oboki, K. et al. IL-33 is a crucial amplifier of innate rather than acquired immunity. *Proc. Natl Acad. Sci. USA* **107**, 18581–18586 (2010).
37. Polumuri, S. K. et al. Transcriptional regulation of murine IL-33 by TLR and non-TLR agonists. *J. Immunol.* **189**, 50–60 (2012).
38. Bonnelykke, K. et al. A genome-wide association study identifies *CDHR3* as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat. Genet.* **46**, 51–55 (2014).
39. Coggins, C. R. A review of chronic inhalation studies with mainstream cigarette smoke in rats and mice. *Toxicol. Pathol.* **26**, 307–314 (1998).
40. Ricciaolo, F. L. M. et al. Expression of vascular remodelling markers in relation to bradykinin receptors in asthma and COPD. *Thorax* **68**, 803–811 (2013).
41. Ricciardolo, F. L. M., Folkerts, G., Folino, A. & Mognetti, B. Bradykinin in asthma: modulation of airway inflammation and remodelling. *Eur. J. Pharmacol.* **827**, 181–188 (2018).
42. Ricciardolo, F. L. M. et al. Bronchoconstriction induced by citric acid inhalation in guinea pigs. *Am. J. Respir. Crit. Care Med.* **159**, 557–562 (1999).
43. Christiansen, S. C. et al. Elevation of tissue kallikrein and kinin in the airways of asthmatic subjects after endobronchial allergen challenge. *Am. Rev. Respir. Dis.* **145**, 900–905 (1992).
44. Wark, P. A. B. et al. Asthmatic bronchial epithelial cells have a deficient innate immune response to infection with rhinovirus. *J. Exp. Med.* **201**, 937–947 (2005).
45. Zhu, J. et al. Bronchial mucosal IFN-α/β and pattern recognition receptor expression in patients with experimental rhinovirus-induced asthma exacerbations. *J. Allergy Clin. Immunol.* **143**, 114–125 (2019).
46. Patel, D. A. et al. Interferon response and respiratory virus control are preserved in bronchial epithelial cells in asthma. *J. Allergy Clin. Immunol.* **134**, 1402–1412 (2014).
47. Altman, M. C. et al. Interferon response to respiratory syncytial virus by bronchial epithelium from children with asthma is inversely correlated with pulmonary function. *J. Allergy Clin. Immunol.* **142**, 451–459 (2018).
48. Durrani, S. R., Viswanathan, R. K. & Busse, W. W. What effect does asthma treatment have on airway remodeling? Current perspectives. *J. Allergy Clin. Immunol.* **128**, 439–448 (2011).
49. Lopez-Guisa, J. M. et al. Airway epithelial cells from asthmatic children differentially express proremodeling factors. *J. Allergy Clin. Immunol.* **129**, 990–997 (2012).
50. Mihaylova, V. T. et al. Regional differences in airway epithelial cells reveal tradeoff between defense against oxidative stress and defense against rhinovirus. *Cell Rep.* **24**, 3000–3007 (2018).

Acknowledgements

This project has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, US Department of Health and Human Services, under contract numbers 1UM1AI114271 awarded to W.W.B. and UM2AI117870 awarded to Rho, Inc. Additional support comes from CTSA UL1TR000150 and UL1TR001422 to J.A.P., 5UL1TR001425 to G.K.K.H., NCRR/NIH UL1TR000451 to M.A.G. and R.S.G., CTSI 1UL1TR001430 to G.T.O., CCTSI UL1TR001082 to A.H.L., 5UM1AI114271 to W.W.B. and J.E.G., NCATS/NIH UL1 TR001876 to S.J.T., and UL1TR002345 to L.B.B. We are grateful to all participants and their families who took part in this study. We thank all of the investigators and staff of the National Institutes of Allergy and Infectious Diseases Inner City Asthma Consortium. We thank the Benaroya Research Institute Genomics and Bioinformatics cores for assistance with sample handling, data generation, and RNA sequencing alignment. We would like to thank P. Woodruff, J. Boyce, and S. Durham for assistance with methodological development as well as advice and discussion.

Author contributions

M.C.A. contributed to study design, RNA-seq data generation, and data analysis and wrote the manuscript. M.A.G. contributed to study design, cell enumeration, and sample collection. E.W. contributed to study design and data analysis. D.C.B. contributed to study design and data analysis. B.S. contributed to cell enumeration. A.H.L. contributed to study design and sample collection. B.J. contributed to data analysis. R.S.G. contributed to sample collection. G.T.O. contributed to sample collection. J.A.P. contributed to sample collection. C.M.K. contributed to sample collection. G.K.K.H. contributed to sample collection. E.M.Z. contributed to sample collection. C.C.J. contributed to sample collection. S.J.T. contributed to sample collection. M.K. contributed to sample collection. L.B.B. contributed to sample collection. A.B. contributed to study design and sample collection. S.M.S. contributed to study coordination. S.P. contributed to RNA-seq data generation. J.E.G. contributed to virology. P.J.G. contributed to study design and study coordination. L.M.W. contributed to study coordination. A.T. contributed to study design and study coordination. W.W.B. contributed to study design, study coordination, and writing of the manuscript. D.J.J. contributed to study design, study coordination, virology, and writing of the manuscript.

Competing interests

M.A.G. reports consulting fees from the American Academy of Allergy, Asthma, and Immunology and the American Academy of Pediatrics. A.H.L. reports consulting fees from Merck Sharp & Dohme and reports data-monitoring committee membership for an asthma study conducted by GlaxoSmithKline. R.S.G. reports employment as a special government employee with the Center for Biologics Evaluation and Research and consulting fees from the Consulting Massachusetts Medical Society. G.T.O. reports

consulting fees from AstraZeneca and reports a grant from Janssen Pharmaceuticals paid to his employing institution. J.A.P. reports provision of study drugs from GlaxoSmithKline, Teva, Merck, Boehringer-Ingelheim, and Genentech/Novartis for research studies outside of the scope of the submitted work. C.M.K. reports consulting fees from GlaxoSmithKline. E.M.Z. reports consulting fees from Wayne State University. S.J.T. reports consulting fees from Novartis, grants from PCORI, the Fight for Children Foundation, EJF Philanthropies, and NIH/NHLBI, and royalties from Uptodate. M.K. reports consulting fees from Novartis. L.B.B. reports consulting fees from Aerocrine, GlaxoSmithKline, Genentech/Novartis, Merck, Cephalon, DBV Technologies, Teva, Boehringer-Ingelheim, AstraZeneca, WebMD/Medscape, Sanofi, Vectura, and Circassia. J.E.G. reports consulting fees from Janssen, Regeneron, and PReP Biosciences and travel expenses from Boehringer-Ingelheim. W.W.B. reports consulting fees from Boston Scientific, ICON, Novartis, GlaxoSmithKline, Genentech, Roche, Boehringer-Ingelheim, Sanofi Genzyme, AstraZeneca, Teva, 3M, PrEPBiopharm, Circassia, Regeneron, Peptinnovate, and Elsevier. D.J.J. reports consulting fees from Novartis, GlaxoSmithKline, Boehringer-Ingelheim, Pfizer, Commense, and Vectura and a grant from NIH/NHLBI. M.C.A., D.C.B., E.W., B.S., B.J., G.K.K.H., C.C.J., A.B., S.M.S., S.P., P.J.G., L.M.W., and A.T. have nothing to disclose.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41590-019-0347-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.C.A.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims inpublished maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Study design and eligibility. This was a prospective, longitudinal case–control study designed to identify changes in gene transcription during cold-associated asthma exacerbations in children. Participants were recruited across nine inner-city sites in Boston, Chicago, Cincinnati, Dallas, Denver, Detroit, New York, St. Louis, and Washington, DC. An individual was eligible for enrollment if he or she was 6 to 17 years of age; was diagnosed with asthma by a clinician more than 1 year prior to recruitment; had at least two asthma exacerbations in the prior year (defined as a requirement for systemic corticosteroids and/or hospitalization); was treated with at least 250 μg of fluticasone delivered by one puff twice daily or its equivalent for those aged 6 to 11 years or treated with at least 250/50 μg of Advair delivered by one puff twice daily or its equivalent for those aged 12 years and older; had a peripheral blood eosinophil concentration of ≥150 cells/mm^3; was a nonsmoker; and had a parent or legal guardian who signed written informed consent and, if applicable, signed the assent form, as per central institutional review board (IRB) guidelines.

An individual was not eligible for enrollment if the individual was pregnant or lactating; was receiving treatment with anti-IgE therapy or had received anti-IgE therapy in the previous 3 months; was receiving immunotherapy; had clinically significant abnormalities on complete blood count; was treated with systemic corticosteroids for any medical condition including an asthma exacerbation within the previous 2 weeks; had a cold within the previous 7 d; was participating in an asthma-related pharmaceutical study or intervention study or had participated in such a study within the previous 4 weeks; required greater than 500 μg of fluticasone plus a long-acting beta-agonist delivered by one puff twice daily or its equivalent and/or used oral corticosteroids daily or every other day; had any medical illnesses that in the opinion of the investigators would increase the risk incurred by participating in the study, interfere with measured outcomes, or interfere with performance of study procedures; had concurrent medical problems that required systemic corticosteroids or other immunomodulators; was diagnosed with cancer or had a history of cancer; had plans to move from the area; did not have a primary caretaker that spoke English or Spanish; was a foster child; would not allow management of their asthma for the duration of the study; was not able to perform pulmonary function tests; had known hypersensitivity to any medications used for asthma treatment; or had a life-threatening asthma exacerbation in the previous 2 years.

Ethical compliance. The study complied with all relevant ethical regulations. The study was approved by a central IRB (Western IRB), and a parent or legal guardian for each child signed written informed consent before study procedures were completed. The study is registered on ClinicalTrials.gov under NCT02502890.

Study visits. Participants meeting eligibility criteria attended a screening and enrollment visit (visit 0). Participants were then followed prospectively for up to two cold events or approximately 6 months after visit 0, whichever occurred first. Participants who reported cold symptoms returned to the clinic for the following visits: visit 1a, which occurred within 3 d of the onset of cold symptoms, and visit 1b, which occurred 4 to 6 d after the onset of cold symptoms. Participants who reported a second set of cold symptoms (occurring at least 2 weeks after the onset of the first cold symptoms and at least 2 weeks after the final dose of systemic corticosteroids was administered for an asthma exacerbation) had additional visits, visit 2a and visit 2b, which were identical in nature to visit 1a and visit 1b. Visits were defined as an exacerbation event (Ex$^+$) if the participant was treated by a physician with systemic corticosteroids in the 10 d following onset of cold symptoms; otherwise, they were defined as a cold without exacerbation (Ex$^-$). Criteria for the initiation of systemic corticosteroids were as previously specified in Inner-City Asthma Consortium trials[4]. Participants were treated with systemic corticosteroids if albuterol was needed by inhaler or nebulization for more than six individual treatments over 24 h; moderate-to-severe wheeze, cough, shortness of breath, and/or chest tightness or pain occurred for at least 5 of the preceding 7 d; symptoms of wheeze, cough, shortness of breath, and/or chest tightness or pain were severe enough to place a participant in his or her 'red zone' on their asthma action plan[4] and did not significantly improve after three doses of albuterol; there was an unscheduled visit for acute asthma care requiring repeated doses of albuterol (to a clinician's office, urgent care, or emergency department); or hospitalization was needed for asthma. Additional visits included visit 3 (3 months after visit 0) for medication management and a final study visit, visit 4, which occurred approximately 2 weeks after the second set of cold visits (2 weeks after visit 2b) or 6 months after visit 0 if the participant did not have two sets of cold visits.

Throughout the study, an internet-based asthma and cold symptom diary was accessed by each participant by using a handheld device provided at enrollment. The diary asked participants to assess cold and asthma symptoms on a four-point visual scale. Participants were asked to complete the diary on a daily basis, and if symptoms of a cold or asthma exacerbation were reported a participant was contacted on the same day or the next weekday by study coordinators to schedule follow-up visits 1a and 1b or visits 2a and 2b. Diary compliance was monitored throughout the study, weekly reminders were provided, and study coordinators contacted participants directly to assess symptoms if compliance was low.

At visits 0, 3, and 4, all participants received asthma step care according to the Guidelines for the Diagnosis and Management of Asthma Expert Panel Report 3 (EPR3)[51]. Treatment was supervised by a study clinician. The study clinician prescribed initial, appropriate concomitant asthma therapies and provided a written asthma action plan for each participant at each visit. The regimen was adjusted as necessary throughout the duration of the study on the basis of the principles in the EPR3 guidelines.

Clinical evaluations conducted at each visit included assessment of the following: vital signs and growth parameters (height, weight, pulse rate, temperature, and blood pressure); medical history and physical examinations; exhaled nitric oxide according to American Thoracic Society guidelines; pulmonary function according to American Thoracic Society and European Respiratory Society guidelines; and a questionnaire, including information on demographics, asthma and allergy history and medication use, asthma exacerbation history, asthma symptoms and healthcare utilization, cold symptoms, concomitant medications, and adverse events.

Allergen skin testing was performed at visit 3 or 4 by using the prick technique with the GreerPick system. Allergens that were assessed included mouse, dog, cat, rat, dust mite, cockroach (American and German), *Alternaria*, *Aspergillus*, ragweed mix, Eastern 8 tree mix, K-O-T grass mix, Bermuda grass, and/or juniper (for some sites only). Skin testing was performed on participants after at least 5 d in which they did not use antihistamines.

Sample collection. *Blood.* Blood samples were collected via peripheral venipuncture at visits 0, 1a, 2a, 1b, and 2b to collect RNA (Tempus Blood RNA tubes, Thermo Fisher Scientific) and to assess complete blood counts with automated differentials. RNA samples were vortexed for 10 s and then frozen at −20 °C until shipment to a central processing facility. Blood differentials were processed by each site's CLIA-approved lab facility through standard methods.

Nasal blow. Nasal mucus samples were obtained at visits 0, 1a, and 2a for respiratory virus assessment as previously described[52,53]. Briefly, participants were instructed to exhale forcefully through the nose into a tissue to clear excess mucus. Holding one nostril shut, saline was sprayed into the other nostril and the participant then blew from that nostril into a collection 'baggie'. The procedure was repeated with the opposite nostril while using the same baggie. M4RT medium (Thermo Fisher Scientific) was added to the collection baggie, and the contents were mixed and stored at −80 °C until shipment.

Nasal lavage. Nasal lavage samples were collected at visits 0, 1a, 1b, 2a, and 2b. Nasal lavage was collected immediately after a nasal blow procedure as follows. Holding their head level, the participant squeezed sterile sodium bicarbonate–buffered normal saline from a 240-ml sinus rinse bottle (NeilMed Sinus Rinse) into the right nostril while a 50-ml Falcon collection tube (tube 1) was held under the left nostril, until 15–20 ml of fluid was collected from the left nostril. Then, occluding the right nostril, the participant gently blew the left nostril several times into a separate 50-ml collection tube (tube 2). Next, while the participant held their head backward at an approximately 45° angle, 5–10 ml of saline from the sinus rinse bottle was squeezed into the right nostril and held in the nose for 2–3 s, after which the participant tilted their head forward and allowed the saline to drip into collection tube 2. This procedure was repeated for the left nostril while using a third collection tube (tube 3) and tube 2. The entire procedure was repeated for nostrils. Samples were kept on ice during and after collection.

After collection, each tube was vortexed for 10 s and consolidated. Samples were centrifuged at 1,300*g* for 10 min at 4 °C. All but ~10 ml of supernatant was removed and discarded. Cell pellets were resuspended in the remaining supernatant. The resuspended pellets were passed through a 100-micron strainer (MACS SmartStrainer). The strained fluid was gently mixed by pipette to homogenize the sample. 1.3 ml was removed and set aside for slide generation, and the remainder was centrifuged at 1,300*g* for 10 min at 4 °C, the supernatant was removed, and 1 ml of RNAprotect Saliva Reagent (Qiagen) was added and the pellet resuspended. Samples were kept on ice and/or at 4 °C during processing.

Urine. Urine was collected at each visit for urine pregnancy testing (for females who had reached menarche) and for urine cotinine testing to measure exposure to environmental tobacco smoke.

Virology. Nasal blow samples were used to assess respiratory viruses by multiplex PCR and partial sequencing to identify RV species and types as previously described[52,53].

Nasal cell differentials. The reserved 1.3 ml of nasal material was used to generate six slides with a Cytospin 4 cytocentrifuge (Thermo Fisher Scientific). Two slides were generated with 500 μl each, two slides with 100 μl each, and two slides with 10 μl each, following the manufacturer's protocol. Slides were air dried for 10 min and then fixed and stained with a Hema3 Stain Set (Thermo Fisher Scientific) according to the manufacturer's protocol.

Nasal cell differentials were performed by manual review of six Cytospin slides per sample, followed by manual read of the slide with the optimal cell

density to distinguish among neutrophils, lymphocytes, macrophages, eosinophils, respiratory epithelium cells, and squamous cells. The following previously established[54] morphological criteria were used to identify these six cell types in the differential cell analysis: (i) neutrophils: segmented bi-, tri-, or multilobed nucleus, pale cytoplasmic staining, and indistinct cytoplasmic borders; (ii) lymphocytes: high nuclear-to-cytoplasmic ratio and small size in relation to macrophages; (iii) macrophages: vacuolated cytoplasm, round shape, variable cell size, and often eccentric nuclei; (iv) eosinophils: cytoplasm filled with intensely pink-stained granules and bi-, tri-, or multilobed nucleus; (v) respiratory epithelium cells: columnar or variable shape, relatively darker cytoplasmic staining as compared to squamous cells, and/or the presence of cilia and/or a terminal nucleus; and (vi) squamous cells: large size, high cytoplasmic-to-nuclear ratio, and bland cytoplasmic staining. Two sequential counts were performed on each sample. All slides were read by a single experienced individual who was blinded to clinical and outcome variables.

RNA-seq library preparation and sequencing. Total RNA was isolated from nasal cell pellets in RNAprotect Saliva Reagent (Qiagen). Samples were centrifuged for 10 min at 10,000*g*, supernatant was removed, and the cell pellet was resuspended in 350 µl of RLT buffer with 1% β-mercaptoethanol. Samples were vortexed for 1–2 s, sonicated for 30 s, spun through a QIAshredder column (Qiagen), and then extracted with an RNeasy MinElute spin column (Qiagen) following the manufacturer's protocol.

Total RNA was isolated from whole-blood lysate by using the MagMAX for Stabilized Blood Tubes RNA isolation Kit for Tempus Blood RNA Tubes (Thermo Fisher Scientific) following the manufacturer's protocol. Globin mRNA was depleted by using GLOBINclear (Thermo Fisher Scientific).

RNA quality was assessed by RNA electrophoresis (Agilent) and NanoDrop 1000 (NanoDrop Products, Thermo Fisher Scientific). Sequencing libraries were constructed from total RNA with TruSeq RNA Sample Preparation Kit v2 (Illumina) and clustered onto a flow cell by using the cBOT amplification system with the HiSeq SR v4 Cluster Kit (Illumina). Single-read sequencing was carried out on a HiSeq 2500 sequencer (Illumina), by using the HiSeq SBS v4 kit to generate 58-base reads, with a target of approximately 10 million reads per sample. Samples were sequenced in four batches and repeated samples were used across batches to allow assessment for batch effects. Resulting bcl files were deconvoluted and converted to fastq format by using Casava from Illumina. fastq files were aligned to the Ensembl version of the human genome (GRCh38; GenBank assembly GCA_000001405.15) with TopHat (version 1.4.1)[55]. The single-paired flag was set to 'single', while all other TopHat parameters were set to default. HTSeq-count[56] was used to generate gene counts with the mode set to 'Intersection (nonempty)' and the minimum alignment quality set to 0; parameters were otherwise set to default. In quality-control analysis, samples were kept that had >0.3 million counts, >0.66 percent alignment, and median CV coverage of <1.5. Genes were filtered to include those that had a trimmed mean of M value (TMM) normalization count of at least 1 in at least 10% of libraries and were classified as protein coding with BioMart[57]. Each of the four sample batches was independently normalized within the batch via TMM normalization and values were transformed to $\log_2$ (counts per million) along with observations level weights by using voomWithQualityWeights from the limma R package[58,59]. Batch processing did not contribute variability in gene expression due to batch effects, so the replicate with the highest percentage alignment was included for analysis and the other replicates were removed. The final dataset included 374 nasal samples composed of 13,672 genes and 387 blood samples composed of 13,316 genes.

Module definition and validation. Cell-type-associated gene coexpression modules were generated by using samples collected from the first fall enrollment season (first two batches of processed samples), which included 42 participants capturing 59 cold events (146 nasal lavage and 176 peripheral blood samples).

Cell type deconvolution was performed by assigning genes to cell types on the basis of statistically significant positive Pearson correlation between gene expression and cell differentials according to CellCODE[60]. Genes without a cell type assignment were considered to not be cell type specific. For the nasal samples, of the 13,672 unique protein-coding genes, cell type deconvolution associated 1,525 genes with neutrophils, 1,192 genes with lymphocytes, 788 genes with macrophages, 672 genes with eosinophils, 241 genes with epithelium cells, and 1,058 gene with squamous cells. 455 genes were associated with two or more cell types, and 8,744 genes were not associated with a specific cell type. For blood samples, of the 13,316 genes, cell type deconvolution associated 2,069 genes with neutrophils, 1,699 genes with lymphocytes, 127 genes with monocytes, 119 genes with eosinophils, and no genes with basophils. 59 genes were associated with two or more cell types, and 9,360 genes were not associated with a specific cell type.

After genes were either assigned to a cell type or left unassigned, WGCNA[61] was run on each separate cell type and the unassigned genes within each sample type, giving a total of 52 nasal modules and 42 blood modules. The parameters used were minCorKME = 0.7 and minKMEtoStay = 0.5, indicating that a module must have a block of genes with Pearson correlation of at least 0.7 to be considered a module and all genes in a module must have a Pearson correlation of at least 0.5 with the module's eigengene. Module-level expression values were calculated by using the geometric mean for the genes assigned to a module.

Coherence of these modules was assessed at completion of the study across the entire sample set spanning both years of study (374 nasal samples and 387 blood samples). We calculated gene–cell type assignments, module–cell type Pearson correlations, and intramodular Pearson correlations and compared these values to the values from module generation. These values were consistent with the levels seen in defining modules and were highly significant (Supplementary Fig. 7a), demonstrating the stability and reproducibility of this modular repertoire for the population (severe asthmatics) and variables (baseline, upper-respiratory illness, and exacerbation) of interest. Final cell assignment for each module was based on significant Pearson correlation of the module and cell differentials in the full dataset. Seven nasal modules were associated with neutrophils, five were associated with lymphocytes, five were associated with macrophages, five were associated with eosinophils, eight were associated with epithelium cells, and eight were associated with squamous cells. Five blood modules were associated with neutrophils, seven were associated with lymphocytes, two were associated with monocytes, and two were associated with eosinophils. All analyses were run in R.

Pathway analysis. The biological function of modules was investigated by using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8[62,63]. This database uses a modified Fisher's exact test to identify specific biological or functional categories that are over-represented in gene sets in comparison with a reference set (the default human genome was used as the reference set). Each module was submitted, and all available enrichment categories were downloaded and ranked according to FDR. Modules were considered to have significant enrichment if one or more of the DAVID default categories showed enrichment with FDR < 0.05. 87% of modules showed enrichment for one or more DAVID category terms. Functional annotation clusters were also generated by using all default categories in DAVID.

STRING version 10.5[64,65], which is a database of known and predicted protein–protein interactions, was used to determine the interaction networks for each module. All STRING interaction sources were used, and a minimum interaction score or 0.15 was required for a gene to be included within a network. An interaction network was significant if its protein–protein interaction enrichment *P* value was <0.05, meaning it had more interactions than expected from a set of proteins of the same size drawn randomly from the human genome. Cytoscape version 3.5.1[66] was used to draw interaction networks according to a prefuse force-directed layout by using the combined interaction score exported from STRING. The size of each gene was made proportional to the number of interactions in the network as a percentage of the maximum number of interactions for any node in that network. Unconnected nodes were excluded. A summary annotation of each module was derived from manual inspection of a module's cell type assignment, functional enrichment, and interaction network.

Bioinformatic analyses. To determine primary sources of variability in module expression, principal-component analysis was run and the first five principal components were correlated (Pearson) to clinical and demographic variables to determine the most important sources of variability in the data (Supplementary Fig. 3). This demonstrated that cell percentages, presence or absence of virus, and visit were each associated with much of the variability in nasal and blood samples. The difference between Ex^+ and Ex^- events was tested by using a weighted linear model (limma in R) and empirical Bayes method[59]. The primary model included terms for exacerbation status, cell percentages, presence or absence of virus, and visit number (1a/2a or 1b/2b) as well as library sequencing depth. Because multiple samples came from each participant, a random effect was included for participant.

The model for nasal samples, with *y* representing each module's expression, was

$$\begin{aligned} y = {} & \beta_0 + \beta_1 \times \text{exacerbation(pos/neg)} + \beta_2 \times \text{lymphocyte(\%)} \\ & + \beta_3 \times \text{eosinophil(\%)} \\ & + \beta_4 \times \text{macrophage(\%)} + \beta_5 \times \text{epithelium(\%)} \\ & + \beta_6 \times \text{squamous(\%)} \\ & + \beta_7 \times \text{viral(pos/neg)} + \beta_8 \times \text{visit(a/b)} + \beta_9 \times \text{libCounts} \\ & + \mu_{\text{individual}} + \varepsilon \end{aligned}$$

The model for blood was

$$\begin{aligned} y = {} & \beta_0 + \beta_1 \times \text{exacerbation(pos/neg)} + \beta_2 \times \text{lymphocyte(\%)} \\ & + \beta_3 \times \text{eosinophil(\%)} \\ & + \beta_4 \times \text{monocyte(\%)} + \beta_5 \times \text{basophil(\%)} \\ & + \beta_6 \times \text{viral(pos/neg)} \\ & + \beta_7 \times \text{visit(a/b)} + \beta_8 \times \text{libCounts} + \mu_{\text{individual}} + \varepsilon \end{aligned}$$

Multiple-testing correction was performed by using the Benjamini–Hochberg method and modules with FDR < 0.05 were considered significant for all comparisons. Other FDR values that were close to significance are also reported in the text.

The same model was used for the paired and unpaired subgroup analyses. For the sensitivity analysis, we performed bootstrapping with random subsets of 80% of participants over 200 iterations.

For the subgroup comparisons (V^+Ex^+, V^+Ex^-, V^-Ex^+, and V^-Ex^-), presence or absence of virus was removed from the model and we ran ANOVA to determine an overall significance value and then pairwise comparisons to determine significant differences across groups. For the corticosteroid analysis, the model included the same variables of cell percentages, presence or absence of virus, visit, and library sequencing depth and then tested contrasts for differences in status (Ex^-, Ex^+ pre-CS, and Ex^+ post-CS) within a visit and across visits. The term for presence or absence of virus was removed in the analysis restricted to V^+ events.

For the longitudinal analysis, we used local polynomial regression fitting (loess) in R, with a degree of smoothing of 1.5 and a polynomial degree of 2. To assess significance, a linear model was fit with a B-spline basis for a polynomial spline with degree 2 for days after cold onset, and ANOVA was run to determine whether exacerbation status was significant.

To study the predictive capabilities of gene expression at the visit 0 baseline, healthy visit, a univariate linear model was built to compare time to exacerbation and each module's expression. Time to exacerbation was defined as the time in days from visit 0 to the first reported exacerbation for each individual, whether or not a participant returned for part a and/or b associated with that particular illness. For participants who did not experience an exacerbation during the study, time to exacerbation was censored to the length of enrollment in the study. Two individuals reported an exacerbation at a later time on the same day as visit 0 and so were excluded from this analysis. Forward stepwise regression was performed with only the significant modules from the univariate analysis in comparison to the time to exacerbation. We selected the model providing the lowest Akaike information criterion that had significant *P* values for each beta coefficient, which was a linear combination of two modules. Samples were divided into quartiles according to their scaled combined expression value for the two modules. A Kaplan–Meier curve showing the exacerbation probability over time was fit to the data splitting individuals according to quartile. The lower three quartiles overlapped and were each significantly different from the top quartile, and they were therefore combined. For comparison, a Kaplan–Meier curve was generated on the basis of FEV_1 predicted (%) split at 80% or split by quartile. For the Kaplan–Meier curves, time was censored at 100 d for display purposes, but the results are similar when using the entire length of follow-up.

All analyses were run in R.

Resources, public data, and analysis code. To present the data and methodology from this study as a resource to the research community, we have made all study data and R code publicly available (see "Data availability" and "Code Availability"). The open source code is also annotated to demonstrate how additional secondary analyses of interest beyond the scope of this manuscript can be performed. Furthermore, we include the R code necessary to generate module expression values for the modular repertoire defined in this manuscript from novel gene expression datasets and run analogous comparisons to those described in this study. Collectively, these resources allow a researcher to readily reproduce the findings of this study, perform secondary analyses within this dataset by using other clinical variables of interest, and apply the modular repertoire defined within this study to new datasets relevant to asthma or other disease states of interest.

Bipartite network. Module gene expression levels and cell differentials from all nasal and blood samples were combined and pairwise Pearson correlations were determined. Because respiratory epithelium and squamous cells were highly correlated in the data, they were summed for this analysis. Hierarchical clustering was performed to create a correlation heat map. A bipartite network was drawn by using Cytoscape version 3.5.1[66] according to a prefuse force-directed layout with all significant pairwise correlation values with FDR < 0.05. Comparisons were overlaid onto the network by coloring significant modules according to fold-change values and shrinking the non-significant modules.

Statistics for clinical data. Categorical variables were compared between Ex^+ and Ex^-events by using generalized linear mixed models assuming a multinomial distribution, with a random effect for participant to account for correlation between values from the same participants. In cases of non-convergence of the model, Fisher's exact test was used and within-participant correlation was not accounted for. Summaries are displayed as percentage (count). Continuous variables were compared between Ex^+ and Ex^- events by using generalized linear mixed models assuming a normal, log-normal, or Poisson distribution as appropriate, with a random effect for participant to account for correlation between values from the same participant. For comparisons made while assuming a log-normal distribution, values of 0 were set to 0.05 prior to model fitting. Summaries are displayed as the median (first quartile, third quartile). All *P* values reported are considered to be descriptive and no adjustments for multiple comparisons were made. SAS version 9.4 software (SAS Institute) was used for all analyses.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The study data and R code to support the findings of this study have been made publicly available. The raw RNA-seq fastq data and minimum information about a high-throughput nucleotide sequencing experiment (MINSEQE) have been deposited to the Gene Expression Omnibus (GEO) with accession numbers GSE115824, GSE115770, and GSE115823. All metadata from the study cohort have been deposited to ImmPort with accession number SDY1387.

Code availability

The R code for all analyses in this manuscript has been annotated and deposited as open-source code in GitHub at https://github.com/BenaroyaResearch/Peds_Asthma_Modules.

References

51. Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma—Summary Report 2007. *J. Allergy Clin. Immunol.* **120**, S94–S138 (2007).
52. Bochkov, Y. A., Grindle, K., Vang, F., Evans, M. D. & Gern, J. E. Improved molecular typing assay for rhinovirus species A, B, and C. *J. Clin. Microbiol.* **52**, 2461–2471 (2014).
53. Kloepfer, K. M. et al. Increased H1N1 infection rate in children with asthma. *Am. J. Respir. Crit. Care Med.* **185**, 1275–1279 (2012).
54. Fahy, J. V. et al. Safety and reproducibility of sputum induction in asthmatic subjects in a multicenter study. *Am. J. Respir. Crit. Care Med.* **163**, 1470–1475 (2001).
55. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
56. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
57. Smedley, D. et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589–W598 (2015).
58. Liu, R. et al. Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res.* **43**, e97 (2015).
59. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
60. Chikina, M., Zaslavsky, E. & Sealfon, S. C. CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. *Bioinformatics* **31**, 1584–1591 (2015).
61. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
62. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
63. Huang, W. D., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
64. Snel, B. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28**, 3442–3444 (2000).
65. Szklarczyk, D. et al. STRINGv10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
66. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

nature research

Corresponding author(s): Matthew C Altman

Last updated by author(s): Jan 23, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	RNA sequencing bcl files were deconvoluted and converted to fastq format using Casava from Illumina. Fastq files were aligned to the Ensembl version of the human genome (GRCh38) (GenBank assembly accession id GCA_000001405.15), using TopHat (version 1.4.1). The single-paired flag was set to "single," while all other TopHat parameters were set to defaults. HTSeq-count was used to generate gene counts with mode as "Intersection (nonempty)" and minimum alignment quality set to 0 and otherwise set to default parameters.
Data analysis	All RNA sequencing analysis was completed using R versions 3.4.3 and 3.4.4 and the packages referenced in the manuscript. Network graphs were generated using Cytoscape version 3.5.1. Clinical data analysis was completed using SAS version 9.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-sequencing data were deposited to the National Center for Biotechnology Information Gene Expression Omnibus with accession number GSE115824. All metadata from the study cohort have been deposited to ImmPort with accession number SDY1387. Data will be made public at the time of publication of the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size was determined using the R package RNASeqPower and confirmed using the R package PROPER. The following assumptions were made: An average RNA sequencing depth of 10 million reads per library, a type I error rate of 0.05 after multiple testing correction using the Benjamini-Hochberg procedure, and a biological CV for samples ranging between 0.20 and 0.45. The biological CV for samples was based on publicly available microarray data, which was then scaled to the extra variation expected in RNA sequencing data based on the publication "Utilizing RNA-Seq data for de novo coexpression network inference." (PMID: 22556371). The target study sample size was selected to allow us to detect effect sizes of at least 1.50 and higher with 80% power. Our enrollment target was set at 220 participants and our exacerbation event target was set at 32 exacerbations. The power calculation was approved by a NIAID statistician independent of the project.

Data exclusions

For data quality control, RNA sequencing data was retained only for samples that had > 0.66% counts aligned to the human genome, a median CV coverage < 1.5, and > 0.3 million counts. Individuals were excluded from the analysis if they did not report any cold events during the study, did not return to clinic for any cold events, or did not have any samples collected during cold events that produced RNA-sequencing data meeting these quality control metrics (see figure 1). From the RNA-sequencing data, genes were filtered to include only those that had a trimmed mean of M values (TMM) normalization count of at least 1 in at least 10% of libraries and were classified as protein coding using BioMart.

Replication

To cross validate our primary outcome results we did each of the following: 1) The cohort was split into a prespecified paired and unpaired analysis according to a predefined statistical analysis plan and the primary outcome calculated for each; 2) A bootstrap sensitivity analysis was performed though iterative subsetting of the full population. Each of these confirmed the primary outcome as presented in the manuscript.

Randomization

For the primary outcome, individuals/samples were organized into groups according to whether they were collected from individuals during cold symptom events that progressed to an asthma exacerbation defined as clinical symptoms that resulted in systemic corticosteroid use within 10 days of cold symptom onset, versus those events that resolved without treatment with systemic corticosteroids. The requirement for systemic corticosteroids and hence the group classification was monitored by an NIAID study manager independent of the study analysts. The group classification was provided to the analysts at the time of data analysis, and the analysis followed a prespecified statistical analysis plan. For the secondary outcome, individuals/samples were further subgrouped into virus associated or not virus associated depending on whether a virus was detected in the associated nasal lavage sample collecting during that cold symptom event. This subgrouping and secondary outcome was also done according to a prespecified statistical analysis plan.

Blinding

Group allocation was provided to the analysts at the time of data analysis and analyses were conducted according to a prespecified statistical analysis plan.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics

An individual was eligible for enrollment if he or she: was 6 to 17 years of age; was diagnosed with asthma by a clinician greater than 1 year prior to recruitment; had at least 2 asthma exacerbation in the prior year (defined as a requirement for systemic corticosteroids and/or hospitalization); was treated with at least fluticasone 250 mcg 1 puff twice daily or its equivalent for those aged 6 to 11 years, or treated with at least Advair 250/50 mcg 1 puff twice daily or its equivalent for those aged 12 years and

	older; had peripheral blood eosinophils ≥ 150 per mm3; was a non-smoker; and had a parent/legal guardian sign a written informed consent and if applicable, signed the assent form, as per central IRB guidelines. Details of clinical characteristics are presented in Table 1.
Recruitment	Participants were recruited across 9 inner city sites in Boston, Chicago, Cincinnati, Dallas, Denver, Detroit, New York, St Louis, and Washington DC. Clinical sites utilized multiple recruitment sources throughout the trial. Sites reviewed lists of participants from previous Inner City Asthma Consortium (ICAC) trials who agreed to future contact as well as lists of potential participants from institutional Emergency Departments, clinics, and EPIC health software databases. Some sites also partnered with asthma and allergy providers in the community. Participants were also recruited through the ICAC Registry (RACR2/ICAC-25). Advertisements were placed in the institutions and around the community to elicit interest in the trial.
Ethics oversight	The study complied with all relevant ethical regulations. The study was approved by a central IRB (Western IRB), and a parent or legal guardian for each child signed written informed consent before completing study procedures.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	The study was registered on clinicaltrials.gov as NCT02502890.
Study protocol	The full trail protocol can be accessed in ImmPort under accession number SDY1387
Data collection	The study sites for data collection are available at clinicaltrials.gov, NCT02502890. The study start was October 2015, recruitment ended October 2016, and study completion was January 2017.
Outcomes	Primary and secondary outcome measures were pre-defined in the study protocol and the statistical analysis plan available in ImmPort under accession number SDY1387.

nature immunology

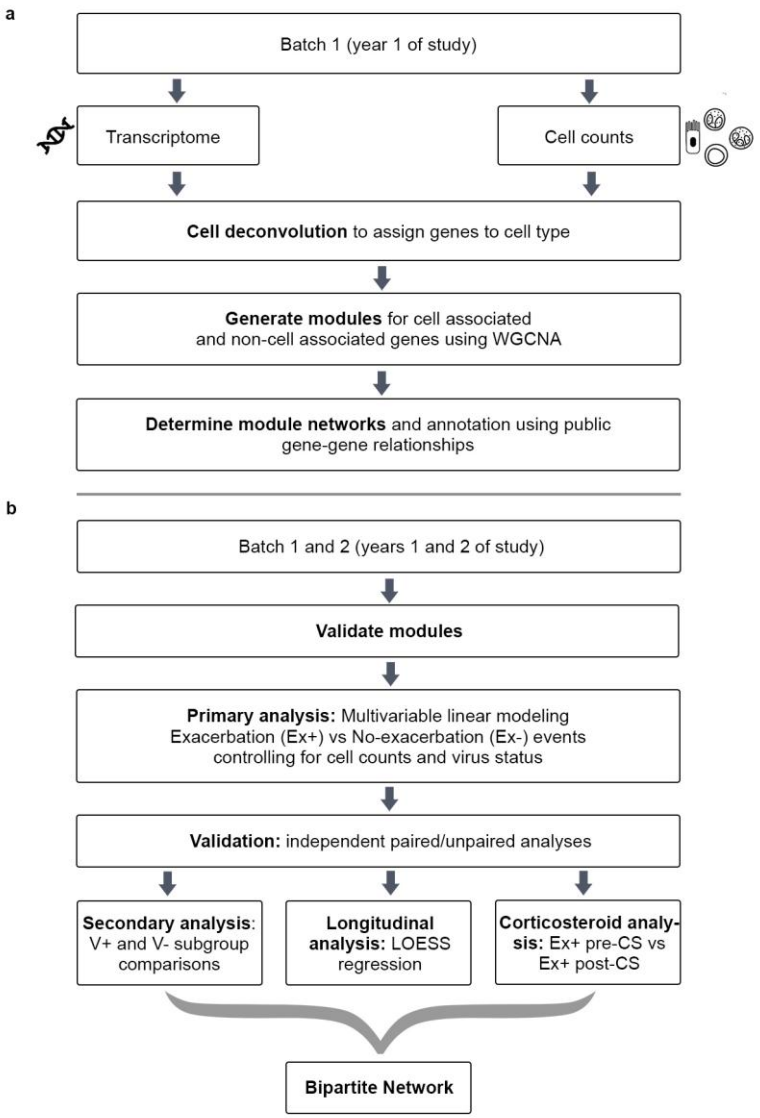
SUPPLEMENTARY INFORMATION
https://doi.org/10.1038/s41590-019-0347-8

In the format provided by the authors and unedited.

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]

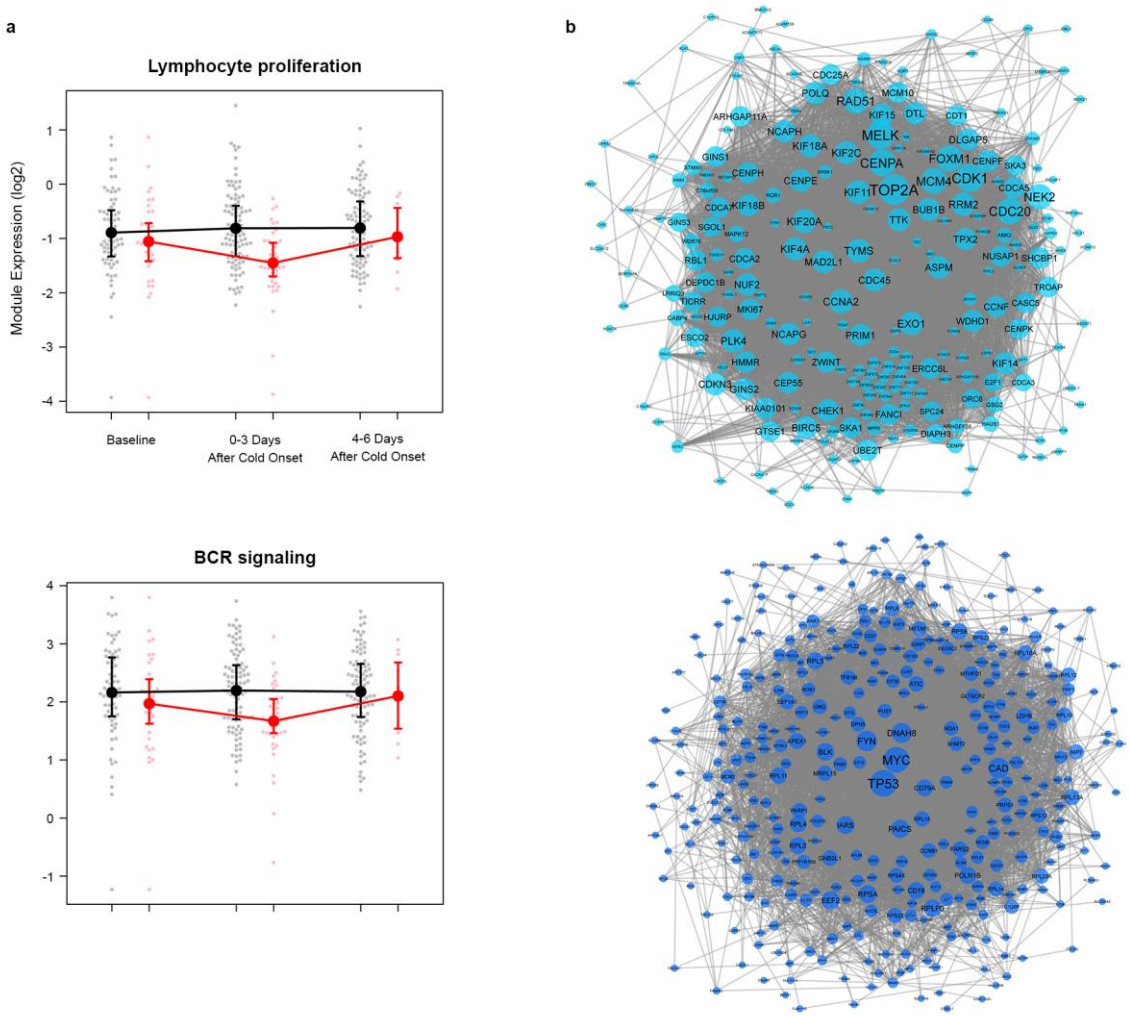
[1]Department of Allergy and Infectious Diseases, University of Washington, Seattle, WA, USA. [2]Systems Immunology Program, Benaroya Research Institute, Seattle, WA, USA. [3]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, USA. [4]Rho, Inc., Chapel Hill, NC, USA. [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

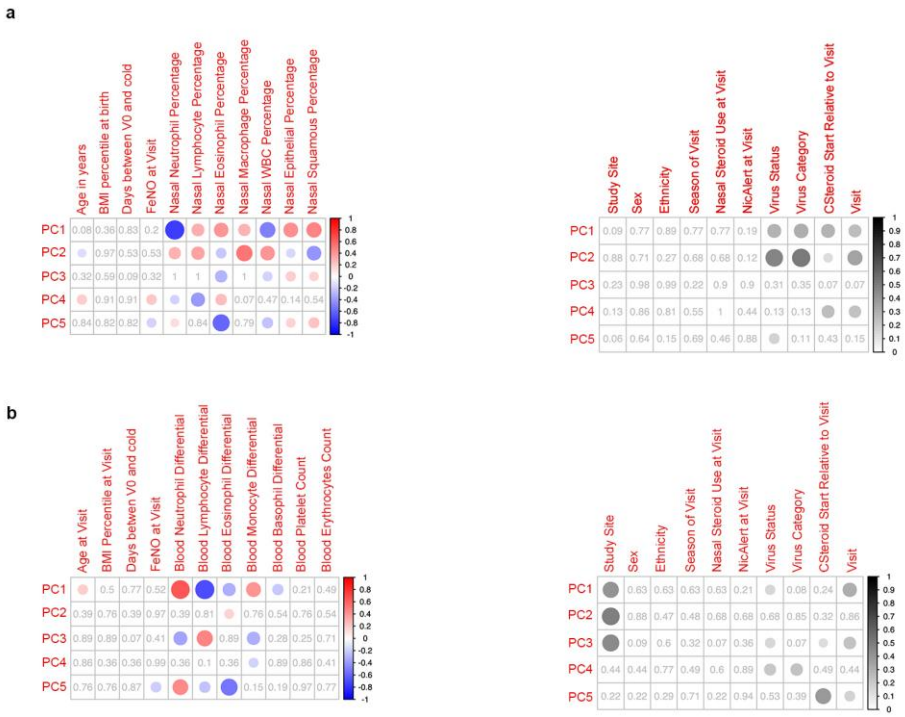
(a) Cell-associated gene coexpression modules were generated using samples collected from the first fall enrollment season, which included 42 participants capturing 59 cold events. Cell deconvolution was performed to assign genes to cell types, and then WGCNA was used to generate coexpression modules. 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.

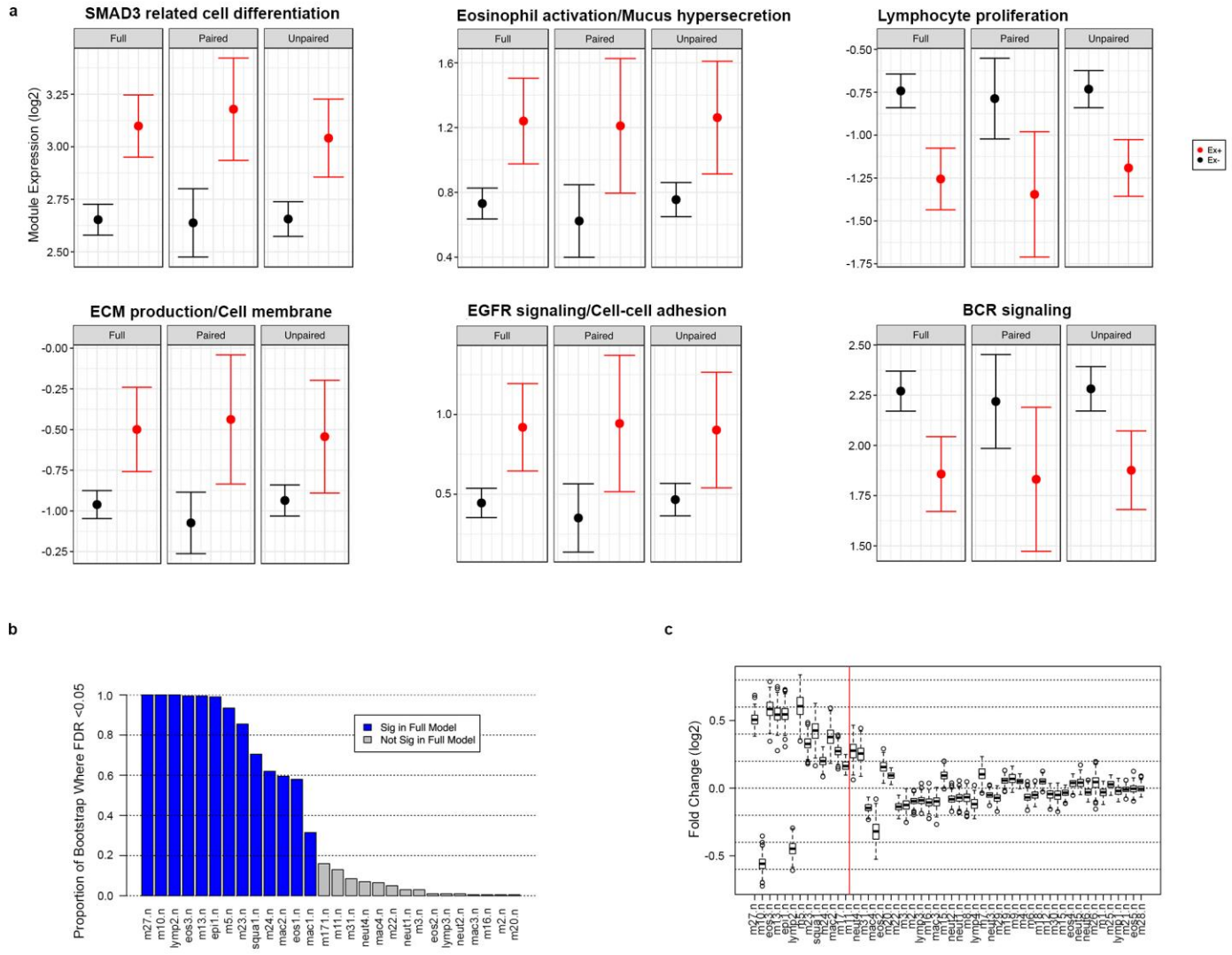
(a) The expression levels of the 2 nasal gene expression modules that decreased in Ex+ events compared to Ex- events. FDR values from top to bottom are 1.38E-05, 4.09E-04. Expression levels represent the log base 2 of the geometric mean of the normalized expression of all genes within the module. Shown 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.

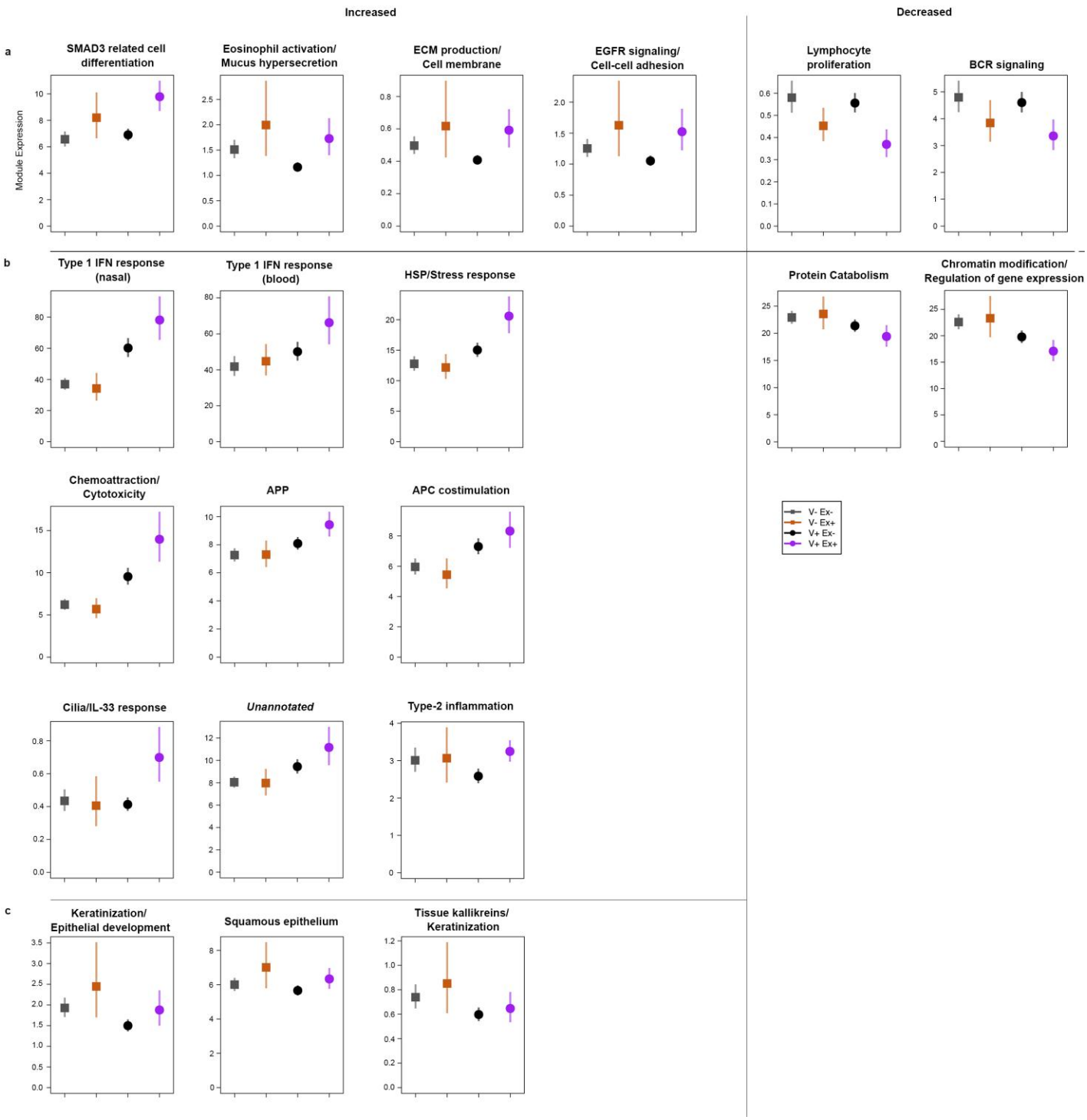
PCA was performed on the expression data and principal components (PCs) were correlated to clinical and demographic variables. To determine significance, Pearson correlation was calculated for continuous variables and ANOVA was run for categorical variables. Significant relationships (FDR<0.05) between 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.



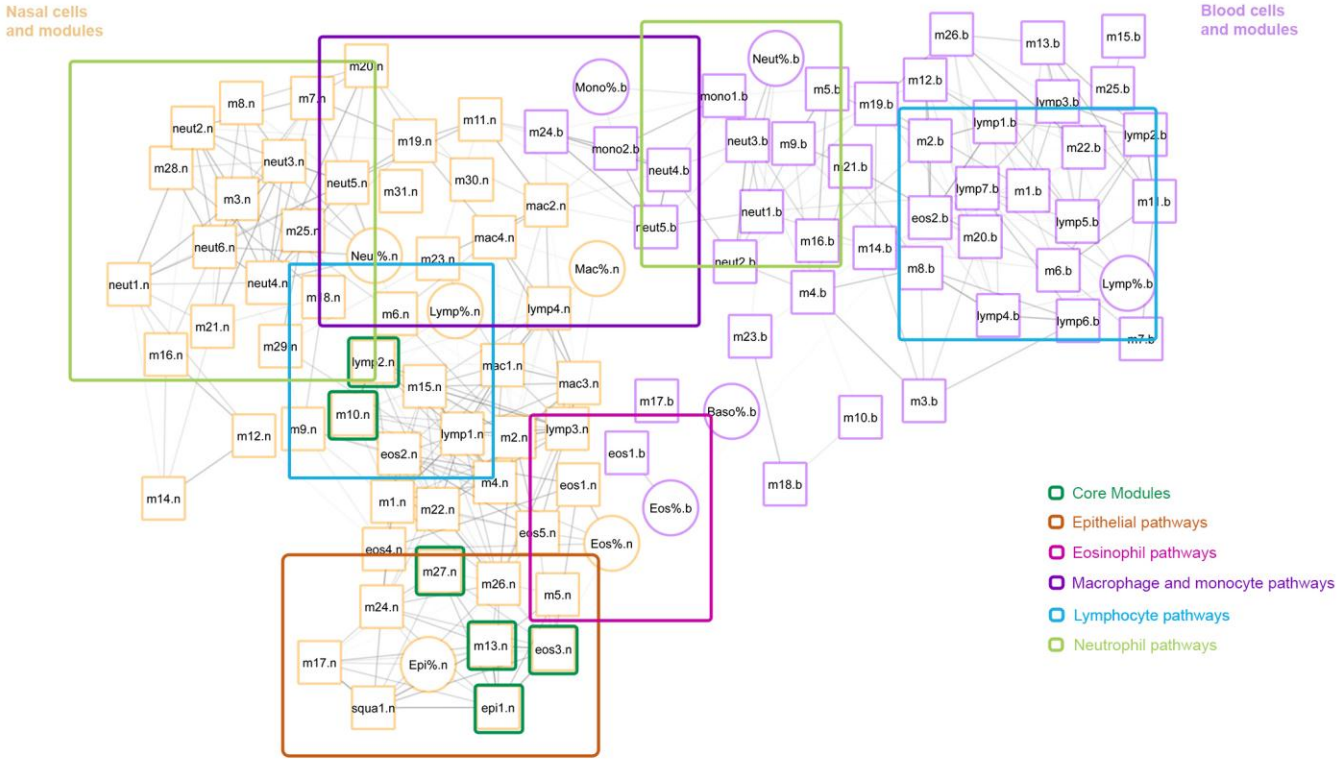
Supplementary Figure 5

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.

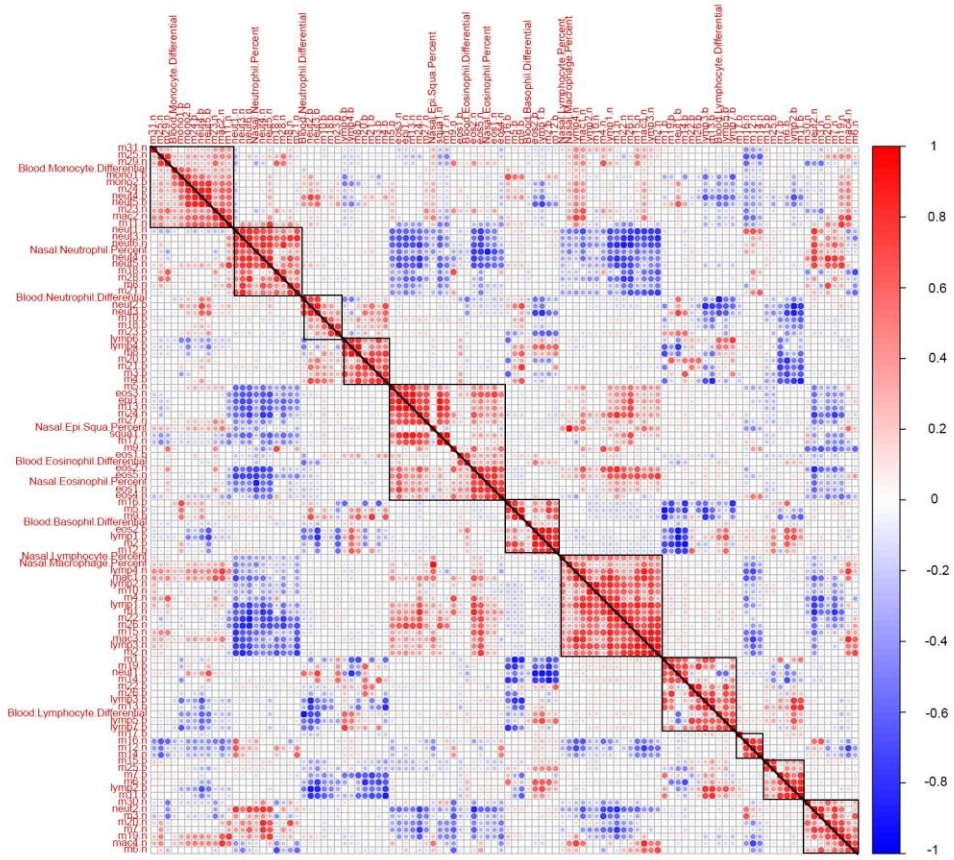
(a) The core exacerbation modules show a similar degree of altered expression in both V+Ex+ (n=33) and V-Ex+ (n=14) events and are significant in each subgroup comparison. (b) V+Ex+ events demonstrate specific upregulation in 9 modules that represent numerous inflammatory pathways and specific downregulation 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

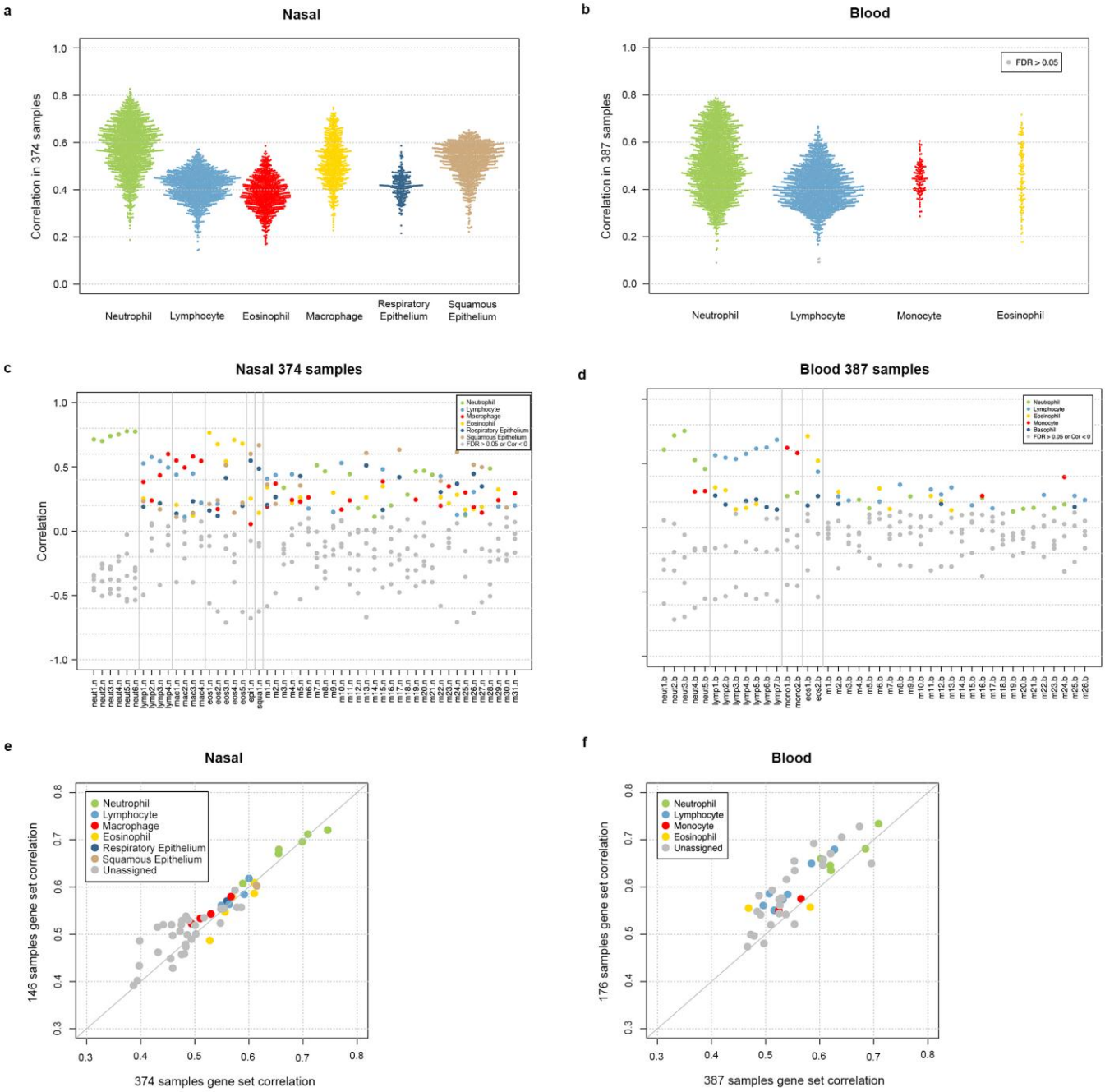
b


Blood.Monocyte.Differential
Nasal.Neutrophil.Percent
Blood.Neutrophil.Differential
Nasal.Epi.Squa.Percent
Blood.Eosinophil.Differential
Nasal.Eosinophil.Percent
Blood.Basophil.Differential
Nasal.Lymphocyte.Percent
Nasal.Macrophage.Percent
Blood.Lymphocyte.Differential
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

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.

nature research

Corresponding author(s): Matthew C Altman

Last updated by author(s): Jan 23, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	RNA sequencing bcl files were deconvoluted and converted to fastq format using Casava from Illumina. Fastq files were aligned to the Ensembl version of the human genome (GRCh38) (GenBank assembly accession id GCA_000001405.15), using TopHat (version 1.4.1). The single-paired flag was set to "single," while all other TopHat parameters were set to defaults. HTSeq-count was used to generate gene counts with mode as "Intersection (nonempty)" and minimum alignment quality set to 0 and otherwise set to default parameters.
Data analysis	All RNA sequencing analysis was completed using R versions 3.4.3 and 3.4.4 and the packages referenced in the manuscript. Network graphs were generated using Cytoscape version 3.5.1. Clinical data analysis was completed using SAS version 9.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-sequencing data were deposited to the National Center for Biotechnology Information Gene Expression Omnibus with accession number GSE115824. All metadata from the study cohort have been deposited to ImmPort with accession number SDY1387. Data will be made public at the time of publication of the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size was determined using the R package RNASeqPower and confirmed using the R package PROPER. The following assumptions were made: An average RNA sequencing depth of 10 million reads per library, a type I error rate of 0.05 after multiple testing correction using the Benjamini-Hochberg procedure, and a biological CV for samples ranging between 0.20 and 0.45. The biological CV for samples was based on publicly available microarray data, which was then scaled to the extra variation expected in RNA sequencing data based on the publication "Utilizing RNA-Seq data for de novo coexpression network inference." (PMID: 22556371). The target study sample size was selected to allow us to detect effect sizes of at least 1.50 and higher with 80% power. Our enrollment target was set at 220 participants and our exacerbation event target was set at 32 exacerbations. The power calculation was approved by a NIAID statistician independent of the project.

Data exclusions

For data quality control, RNA sequencing data was retained only for samples that had > 0.66% counts aligned to the human genome, a median CV coverage < 1.5, and > 0.3 million counts. Individuals were excluded from the analysis if they did not report any cold events during the study, did not return to clinic for any cold events, or did not have any samples collected during cold events that produced RNA-sequencing data meeting these quality control metrics (see figure 1). From the RNA-sequencing data, genes were filtered to include only those that had a trimmed mean of M values (TMM) normalization count of at least 1 in at least 10% of libraries and were classified as protein coding using BioMart.

Replication

To cross validate our primary outcome results we did each of the following: 1) The cohort was split into a prespecified paired and unpaired analysis according to a predefined statistical analysis plan and the primary outcome calculated for each; 2) A bootstrap sensitivity analysis was performed though iterative subsetting of the full population. Each of these confirmed the primary outcome as presented in the manuscript.

Randomization

For the primary outcome, individuals/samples were organized into groups according to whether they were collected from individuals during cold symptom events that progressed to an asthma exacerbation defined as clinical symptoms that resulted in systemic corticosteroid use within 10 days of cold symptom onset, versus those events that resolved without treatment with systemic corticosteroids. The requirement for systemic corticosteroids and hence the group classification was monitored by an NIAID study manager independent of the study analysts. The group classification was provided to the analysts at the time of data analysis, and the analysis followed a prespecified statistical analysis plan. For the secondary outcome, individuals/samples were further subgrouped into virus associated or not virus associated depending on whether a virus was detected in the associated nasal lavage sample collecting during that cold symptom event. This subgrouping and secondary outcome was also done according to a prespecified statistical analysis plan.

Blinding

Group allocation was provided to the analysts at the time of data analysis and analyses were conducted according to a prespecified statistical analysis plan.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics

An individual was eligible for enrollment if he or she: was 6 to 17 years of age; was diagnosed with asthma by a clinician greater than 1 year prior to recruitment; had at least 2 asthma exacerbation in the prior year (defined as a requirement for systemic corticosteroids and/or hospitalization); was treated with at least fluticasone 250 mcg 1 puff twice daily or its equivalent for those aged 6 to 11 years, or treated with at least Advair 250/50 mcg 1 puff twice daily or its equivalent for those aged 12 years and

	older; had peripheral blood eosinophils ≥ 150 per mm3; was a non-smoker; and had a parent/legal guardian sign a written informed consent and if applicable, signed the assent form, as per central IRB guidelines. Details of clinical characteristics are presented in Table 1.
Recruitment	Participants were recruited across 9 inner city sites in Boston, Chicago, Cincinnati, Dallas, Denver, Detroit, New York, St Louis, and Washington DC. Clinical sites utilized multiple recruitment sources throughout the trial. Sites reviewed lists of participants from previous Inner City Asthma Consortium (ICAC) trials who agreed to future contact as well as lists of potential participants from institutional Emergency Departments, clinics, and EPIC health software databases. Some sites also partnered with asthma and allergy providers in the community. Participants were also recruited through the ICAC Registry (RACR2/ICAC-25). Advertisements were placed in the institutions and around the community to elicit interest in the trial.
Ethics oversight	The study complied with all relevant ethical regulations. The study was approved by a central IRB (Western IRB), and a parent or legal guardian for each child signed written informed consent before completing study procedures.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	The study was registered on clinicaltrials.gov as NCT02502890.
Study protocol	The full trail protocol can be accessed in ImmPort under accession number SDY1387
Data collection	The study sites for data collection are available at clinicaltrials.gov, NCT02502890. The study start was October 2015, recruitment ended October 2016, and study completion was January 2017.
Outcomes	Primary and secondary outcome measures were pre-defined in the study protocol and the statistical analysis plan available in ImmPort under accession number SDY1387.