

Letter to the Editor

Cockroach-induced *IL9*, *IL13*, and *IL31* expression and the development of allergic asthma in urban children*To the Editor:*

Type 2 (T2) immune responses are key risk factors in childhood for airway inflammation, bronchial hyperresponsiveness, and development of asthma. Developing sensitization to multiple allergens in the first few years of life is closely associated with the development of T2 inflammation and is especially important in light of its poor prognosis with respect to persistent asthma, increased exacerbations, and low lung function.¹ Although some risk factors for the development of T2 inflammation and asthma are known, the longitudinal patterns of immune development that drive the imbalance toward a T2 immune response in humans are not fully understood. Furthermore, the critical aspects of persistent T2 inflammation that are responsible for the development of allergic asthma as opposed to allergic sensitization without asthma are unknown.

In previous work, we identified a set of 244 genes, including the T2-associated cytokine genes *IL9* and *IL13*, the levels of expression of which were elevated in allergen-stimulated PBMCs in a group of predominantly African-American children aged 2 years who went on to develop multiple aeroallergen sensitizations and asthma by the age of 7 years.² Here we have extended that analysis by repeating RNA sequencing of the PBMC samples collected longitudinally from children at the ages of 2, 5, and 7 years. We utilized the Urban Environment and Childhood Asthma (URECA) birth cohort at high risk of asthma.³ From the 560 newborns enrolled into URECA, we used a nested case-control design to study 60 children who had available longitudinal RNA samples from cockroach (CR) allergen-stimulated PBMCs collected when they were aged 2, 5, and 7 years and who also had clinical outcome data through 7 years of age. Development of asthma at age 7 years was defined on the basis of a prespecified algorithm that included a doctor's diagnosis, lung function criteria, or symptom criteria; sensitization was based on skin testing and serum IgE level testing (for the full details of group classification, see the [Methods](#) section of this article's Online Repository at www.jacionline.org). Specifically, we compared 14 individuals who developed both asthma and sensitization to multiple aeroallergens, including CR (the asthma and sensitization group), with 15 clinically matched individuals who developed asthma without any aeroallergen sensitizations (the asthma-only group), 15 clinically matched individuals who developed CR sensitization without asthma (the sensitization-only group), and 16 clinically matched individuals who did not develop aeroallergen sensitizations or asthma (the neither group) ([Table 1](#)). We utilized sparse partial least squares discriminant analysis (PLS-DA) as an exploratory tool to determine which of 15,872 longitudinally measured genes provided the most robust classification of clinical outcome. This supervised analysis method is a powerful feature selector and classifier that captures the variability of both explanatory and dependent variables while reducing the relationships to a minimal set of genes that differentiate the outcomes (see the [Methods](#) of the Online Repository).

At each year of study (ages 2, 5, and 7) and averaged across all years, the first explanatory variables of the model specifically differentiated the group of individuals with both asthma and sensitization from the other 3 groups ([Fig 1, A](#)). A total of 52 genes contributed to the sparse PLS-DA models in uniquely separating the group with both asthma and sensitization from the other 3 groups either at a single time point and/or averaged across the 3 time points of study (see [Table E1](#) in this article's Online Repository at www.jacionline.org). Specifically, 3 genes, *IL9*, *IL13*, and *IL31*, showed significant contribution to differentiating the group with both asthma and sensitization from the other 3 groups at multiple time points, and the levels of these genes were all elevated in the group with both asthma and sensitization at each time point ([Fig 1, B](#)). The other 49 genes contributed to a single sparse PLS-DA model either at 1 time point or when averaged across the 3 time points of study, although they all also show similar trends across the 3 time points (see [Table E1](#)). These genes include multiple other cytokines/cytokine receptor genes, including *LTA*, *TGFBR2*, *CSF2*, *FASLG*, *CCL25*, *CSF2RB*, *IL12RB2*, *IL15RA*, and *IL17RB* (KEGG_Pathway, Cytokine-cytokine receptor interaction; false discovery rate [FDR] = 4.9E-6), and also show enrichment for genes signaling through JAK-STAT pathways (KEGG_Pathway, JAK-STAT signaling pathway, FDR = 1.9E-3), as well as multiple other established immune response genes, including *GZMB*, *PRF1*, *CASP3*, *ENPP1*, *EOMES*, *GATA1*, and *SOCST1* (GO_BP, Immune system process, FDR = 8.7E-7).

Our results demonstrate throughout childhood a persistent elevation of the expression of specific systemic T2 cytokines and other immune response genes in PBMCs in response to CR allergen stimulation as a prominent feature of the development of allergic sensitization with asthma in a high-risk urban population. We showed that this response is established by 2 years of age and persists through ages 5 and 7 years, during the period of development of multiple aeroallergen sensitizations and clinical asthma. These results demonstrate elevated expression of *IL9*, *IL13*, and *IL31* as independent features that together are associated with the development of asthma with sensitization by age 7 years.

IL-13 is an archetypal T2 cytokine that has already been well established as a contributor to asthma pathogenesis and is involved in bronchial hyperresponsiveness, goblet cell differentiation, and IgE production.⁴ It represents an attractive therapeutic target in established disease,⁵ and our data add weight to the possibility of considering it as an early-life target for prevention or modification of disease onset. IL-9, which is also produced by T_H2 cells as well as by several other types of lymphocytes, can influence multiple aspects of T2 inflammation in relation to asthma and is perhaps best established in promoting mast cell development.⁶ IL-31 plays a well-established role in atopic dermatitis⁷ and has also now been linked to persistent asthma.⁸ Notably, all 3 of these cytokines are induced by IL-4, and they all also signal through JAK-STAT transduction pathways, suggesting that persistent signaling through the IL-4 receptor may be driving our finding. We previously reported that allergen-induced IL-4 secretion from stimulated PBMCs was related to early onset of allergic sensitization in URECA.⁹ The current analysis extends

TABLE I. Participant characteristics

Characteristic	Asthma and sensitization	Asthma only	Sensitization only	Neither	P value*
	n = 14	n = 15	n = 15	n = 16	
Site, no. (%)					.714
Baltimore	5 (35.7%)	3 (20.0%)	2 (13.3%)	6 (37.5%)	
Boston	3 (21.4%)	6 (40.0%)	4 (26.7%)	6 (37.5%)	
New York	3 (21.4%)	3 (20.0%)	3 (20.0%)	2 (12.5%)	
St Louis	3 (21.4%)	3 (20.0%)	6 (40.0%)	2 (12.5%)	
Mother had asthma, no. (%)	6 (42.9%)	10 (66.7%)	8 (53.3%)	8 (50.0%)	.624
Race of child, no. (%)					.572
Black	12 (85.7%)	10 (66.7%)	13 (86.7%)	13 (81.2%)	
Mixed/other	1 (7.14%)	1 (6.67%)	0 (0.00%)	2 (12.5%)	
White	1 (7.14%)	4 (26.7%)	2 (13.3%)	1 (6.25%)	
Sex, no. (%)					.735
Female	3 (21.4%)	4 (26.7%)	3 (20.0%)	6 (37.5%)	
Male	11 (78.6%)	11 (73.3%)	12 (80.0%)	10 (62.5%)	
Cesarean section, no. (%)	5 (35.7%)	5 (33.3%)	5 (33.3%)	2 (12.5%)	.405
Season of birth, no. (%)					.802
April-June	1 (7.14%)	5 (33.3%)	2 (13.3%)	4 (25.0%)	
January-March	2 (14.3%)	3 (20.0%)	4 (26.7%)	3 (18.8%)	
July-September	5 (35.7%)	4 (26.7%)	6 (40.0%)	5 (31.2%)	
October-December	6 (42.9%)	3 (20.0%)	3 (20.0%)	4 (25.0%)	
Mother's age at birth (y), mean $\pm$ SD	24.8 $\pm$ 4.68	24.5 $\pm$ 7.72	24.1 $\pm$ 5.84	23.8 $\pm$ 6.05	.97
Ever breast-fed, no. (%)	2 (14.3%)	2 (13.3%)	4 (26.7%)	2 (12.5%)	.768
Sum of colds in year 2, mean $\pm$ SD	4.50 $\pm$ 4.31	6.40 $\pm$ 3.31	4.27 $\pm$ 2.66	4.81 $\pm$ 2.56	.281
CR bedroom allergen in year 2, median [IQR]	-0.17 [-1.63 to 2.03]	-1.63 [-1.63 to 2.21]	-1.63 [-1.63 to 2.74]	-1.63 [-1.63 to 1.24]	.787
Age of CR sensitization (y), mean $\pm$ SD	3.0 $\pm$ 2.04	4.7 $\pm$ 1.94	N/A	N/A	.027
Eczema reported at any time (in years 1-7), no. (%)	12 (85.7%)	12 (80.0%)	10 (66.7%)	5 (31.2%)	.009

IQR, Interquartile range [25th percentile; 75th percentile]; NA, not available.

*P values were calculated with chi-square tests for categorical variables and Kruskal-Wallis tests for continuous variables.

these findings to link expression of IL-4-related genes beginning in early childhood to persistent allergic asthma.

These 3 cytokines are likely related to the development of allergic asthma and other allergic diseases as part of a broader immune response that includes several other T2 cytokines, their receptors, and their effects on clinically relevant organs, such as the glandular apparatus, sensory nerves, and airway smooth muscle. Although this observational study cannot distinguish causality, it seems likely that early-onset, strong, and sustained multicytokine T2 responses could promote sensitization and influence airway physiology to promote asthma. These patterns of gene expression could be useful as indicators that could be used early in life to stratify risk of development of allergic asthma in minority children—a hypothesis that needs validation in other studies and in other populations. They may also suggest pathways, in particular, IL-4 receptor signaling (or possibly JAK-STAT signaling), that could be directly modulated in an effort to prevent or modify disease development in high-risk individuals in the first years of life.

We are grateful to the entire URECA collaboration, which consists of the following institutions and investigators (principal investigators are indicated by asterisks): Johns Hopkins University School of Medicine, Baltimore, Maryland (R. Wood,* S. Leimenstoll, and R. Spellman); Boston University School of Medicine, Boston, Massachusetts (G. O'Connor,* M. Sandel, R. Cohen, L. Galalis, E. Gjerasi, and N. Gonzalez; Columbia University Medical Center, New York, New York (M. Kattan,* E. Arteago-Solis, A. Bello, Y. Fernandez-Pau, C. Lamm, S. Lovinsky-Desir, D. Perlaza, L. Peters, M. Pierce, C. Sanabia, S. Tsang, A. Valones, N. Whitney, and P. Yaniv); Washington University School of Medicine and St Louis Children's Hospital, St Louis,

Missouri (L. Bacharier,* A. Beigelman, G. Bloomberg, E. Tesson, and A. Freie); National Institute of Allergy and Infectious Diseases, Bethesda, Maryland (P. Gergen, A. Togias, E. Smartt, and C. Czarniecki); University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin (J. Gern, W. Busse,* D. Jackson, C. Sorkness, S. Doyle, R. Burton, P. Heinrich, K. Spring, A. Dresen, and S. Ramratnam); and Rho, Inc, Durham, North Carolina (S. Arbes,* C. Visness,* A. Calatroni, K. Jaffee, M. Yaeger, and H. Mitchell). We are grateful to all the families who have participated in the URECA study.

Petra LeBeau, ScD^{a*}
 Alexandre Lockhart, MS^{a*}
 Alkis Togias, MD^b
 Megan T. Sandel, MD^c
 Jessica D. Gereige, MD^c
 Leonard Bacharier, MD^d
 Stephanie Lovinsky-Desir, MD^e
 Robert A. Wood, MD^f
 Robert James, PhD^a
 William W. Busse, MD^g
 James E. Gern, MD^g
 Matthew C. Altman, MD^{h,i}

for the National Institute of Allergy and Infectious Diseases-Sponsored Inner-City Asthma Consortium

From ^aRho Inc, Durham, NC; ^bthe US National Institute of Allergy and Infectious Diseases; ^cthe Boston University School of Medicine, Boston, Mass; ^dthe Washington University School of Medicine, St Louis, Mo; ^ethe Columbia University College of Physicians and Surgeons, New York, NY; ^fthe Johns Hopkins University School of Medicine, Baltimore, Md; ^gthe University of Wisconsin School of Medicine and Public Health, Madison, Wis; ^hthe University of Washington Department of Medicine, Seattle, Wash; and ⁱthe Benaroya Research Institute Systems Immunology Division, Seattle, Wash. E-mail: maltman@benaroyaresearch.org.

*These authors contributed equally to this work.

Dr Togias' coauthorship on this publication does not constitute endorsement by the US National Institute of Allergy and Infectious Diseases or by any other United States government agency.

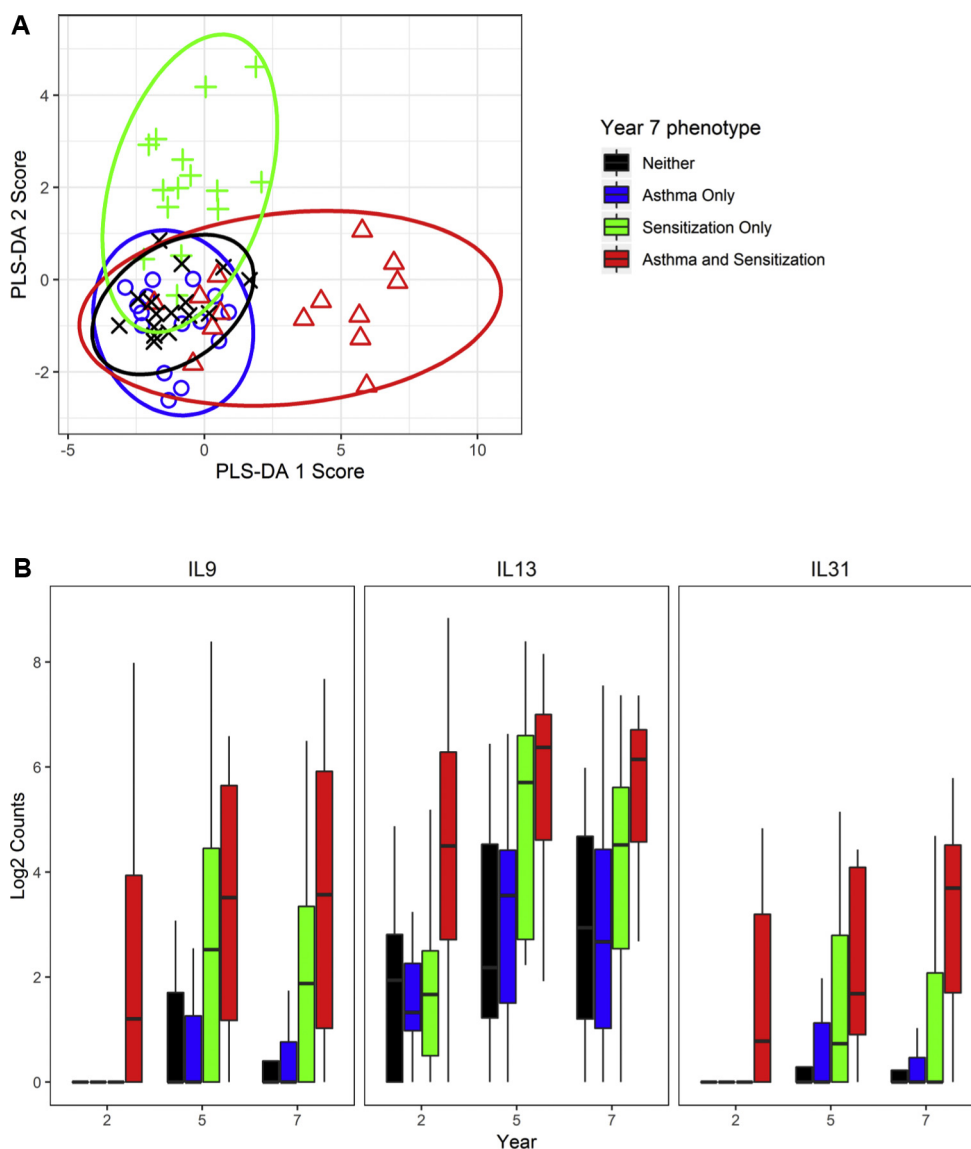


FIG 1. Supervised analysis and feature selection with sparse PLS-DA using the geometric mean gene expression in children at the ages 2, 5, and 7 years. **A**, Sparse PLS-DA sample plot of the first 2 components with 95% CI ellipse plots to highlight the strength of the discrimination between phenotype groups. Each plotted symbol represents a child classified by his or her 7-year phenotype, and each phenotype is represented by a distinct colored symbol. The 15 genes selected on the first sparse PLS-DA component all had positive weight in the linear combination and were highly expressed in the phenotype with both asthma and sensitization. **B**, Boxplots of normalized gene expression counts at years 2, 5, and 7 by year 7 phenotype for *IL9*, *IL13*, and *IL31*.

Supported with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health (under contract numbers NO1-AI-25496, NO1-AI-25482, HHSN272200900052C, HHSN272201000052I, 1UM1AI114271-01, and UM2AI117870), as well as with additional support was provided by the National Center for Research Resources, National Institutes of Health (under grants RR00052, M01RR00533, 1UL1RR025771, M01RR00071, 1UL1RR024156, UL1TR000040, UL1TR001079, and 5UL1RR024992-02). All of the authors, with the exception of A.T., J.D.G., and M.C.A., report grants from the National Institutes of Health/National Institute of Allergy and Infectious Diseases/Division of Allergy, Immunology, and Transplantation during the conduct of the study.

Disclosure of potential conflict of interest: L. Bacharier reports personal fees from GlaxoSmithKline, Genentech/Novartis, DBV Technologies, AstraZeneca, WebMD/Medscape, Sanofi/Regeneron, Vectura, and Circassia, as well as personal fees and nonfinancial support from Merck, Teva, and Boehringer Ingelheim outside the submitted work. J. E. Gern reports consulting fees from Regeneron, consulting fees and stock options from Meissa Vaccines, Inc, and consulting fees from AstraZeneca;

personal fees from PREP Biopharm, Inc, MedImmune, and Ena Pharmaceuticals; and a pending patent for methods of propagating rhinovirus C in previously unsusceptible cell lines and adapted rhinovirus C outside the submitted work. J. D. Gereige reports grants from the National Institutes of Health (T32) outside the submitted work. W. W. Busse reports personal fees from Boston Scientific, Novartis, GlaxoSmithKline, Genentech, Sanofi/Genzyme, AstraZeneca, Teva, Regeneron, and Elsevier outside the submitted work. R. A. Wood reports grants from DNV, Aimmune, Astellas, HAL-Allergy, and Regeneron and royalties outside the submitted work. M. C. Altman reports personal fees from Sanofi-Regeneron outside the submitted work. The rest of the authors declare that they have no relevant conflicts of interest.

REFERENCES

1. Havstad S, Johnson CC, Kim H, Levin AM, Zoratti EM, Joseph CLM, et al. Atopic phenotypes identified with latent class analyses at age 2 years. *J Allergy Clin Immunol* 2014;134:722-7.e2.

2. Altman MC, Whalen E, Togias A, O'Connor GT, Bacharier LB, Bloomberg GR, et al. Allergen-induced activation of natural killer cells represents an early-life immune response in the development of allergic asthma. *J Allergy Clin Immunol* 2018; 142:1856-66.
3. Gern JE, Visness CM, Gergen PJ, Wood RA, Bloomberg GR, O'Connor GT, et al. The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. *BMC Pulm Med* 2009;9:17.
4. Corren J. Role of interleukin-13 in asthma. *Curr Allergy Asthma Rep* 2013;13: 415-20.
5. Castro M, Corren J, Pavord ID, Maspero J, Wenzel S, Rabe KF, et al. Dupilumab efficacy and safety in moderate-to-severe uncontrolled asthma. *N Engl J Med* 2018;378:2486-96.
6. Koch S, Söpel N, Finotto S. Th9 and other IL-9-producing cells in allergic asthma. *Semin Immunopathol* 2017;39:55-68.
7. Ruzicka T, Hanifin JM, Furue M, Pulka G, Mlynarczyk I, Wollenberg A, et al. Anti-interleukin-31 receptor A antibody for atopic dermatitis. *N Engl J Med* 2017;376:826-35.
8. Stott B, Lavender P, Lehmann S, Pennino D, Durham S, Schmidt-Weber CB. Human IL-31 is induced by IL-4 and promotes TH2-driven inflammation. *J Allergy Clin Immunol* 2013;132:446-54.e5.
9. Gern JE, Calatroni A, Jaffee KF, Lynn H, Dresen A, Cruikshank WW, et al. Patterns of immune development in urban preschoolers with recurrent wheeze and/or atopy. *J Allergy Clin Immunol* 2017;140:836-44.

<https://doi.org/10.1016/j.jaci.2021.01.022>

METHODS

URECA population

URECA is a birth cohort study initiated in 2005 in inner-city Baltimore, Boston, New York City, and St Louis; the details of the study design have been described elsewhere.^{E1} In brief, pregnant women aged 18 or older were recruited with selection criteria that included a history of asthma, allergic rhinitis, or eczema in the mother or father. Between February 2005 and March 2007, 1850 families were screened; 776 met eligibility criteria, and 560 newborns were enrolled at birth. Informed consent was obtained from the parent or legal guardian of each infant. Levels of allergen-specific IgE (ImmunoCAP, Phadia, Uppsala, Sweden) for milk, egg, peanut, and German CR were measured yearly until the children had reached the age of 7 years. Levels of specific IgE for dust mites (*Dermatophagoides farinae* and *Dermatophagoides pteronyssinus*), dog, cat, mouse, and *Alternaria* were added for children aged 2 to 7 years. The lower limit of detection for the specific IgEs was 0.10 kU/L, and we manually set all values that were below this to 0.09 kU/L. Prick skin testing for 14 common aeroallergens was performed when the children were 3, 5, and 7 years of age. Allergen sensitization was defined as either a wheal 3 mm in diameter or larger than the saline control on prick skin testing or a specific IgE level of at least 0.35 kU/L. Asthma at age 7 was defined by using a prespecified algorithm that included parent-reported asthma diagnosis by a physician, asthma symptoms, health care use for asthma, use of asthma medications in the previous year, spirometry with reversibility, and bronchial hyperresponsiveness assessed by methacholine inhalation challenge.^{E2}

For this nested case-control study, URECA participants who met the following inclusion criteria were selected for analysis. For the asthma and CR-sensitized group (n = 14), children had to be URECA participants with (1) at least 2 aeroallergen sensitizations at age 3, 5, and/or 7, including sensitization to CR at each age; (2) a diagnosis of asthma at age 7; and (3) PBMC RNA samples collected at ages 2, 5, and 7. For the asthma-only group (n = 15), children had to be URECA participants with (1) no aeroallergen sensitization through age 7 years; (2) a diagnosis of asthma at age 7 years; and (3) PBMC RNA samples collected when they were 2, 5, and 7 years old. For the CR-sensitized group (n = 15), children had to be URECA participants with (1) at least 2 aeroallergen sensitizations at age 3, 5, and/or 7 years, including sensitization to CR at each age; (2) no asthma at age 7 years; and (3) PBMC RNA samples collected when they were 2, 5, and 7 years old. For the neither group (n = 16), children had to be URECA participants with (1) no aeroallergen sensitization through age 7 years; (2) no asthma at age 7 years; and (3) PBMC RNA samples collected when they were 2, 5, and 7 years old. The latter 3 groups were selected to match to the asthma and CR-sensitized group on at least a bivariate level according to the following variables: sex, race, URECA site, and birth season. There were no differences in the 3 components of the asthma diagnosis between the asthma and CR-sensitized group and the asthma-only group.

PBMC processing and stimulation studies

Peripheral blood samples were collected when the children were 2, 5, and 7 years old and processed as previously described.^{E1} In short, at each URECA research center, within 16 hours of blood collection, 1.0×10^6 mononuclear cells were separated and incubated (for 48 hours at 37°C in 5% CO₂) in the presence of German CR extract (10 µg/mL) (Greer Laboratories, Lenoir, NC). Following incubation, the cells were lysed by using TRIzol (Tri-Reagent LS; Fisher Scientific, Waltham, Mass), and maintained at -80°C pending extraction of total cellular RNA.

RNA sequencing

RNA was extracted from the TRIzol following the manufacturer's protocol. RNA quality was assessed by using RNA electrophoresis (Agilent, Santa Clara, Calif). All samples had RNA integrity number values of 7 or higher. Sequencing libraries were constructed from total RNA by using TruSeq RNA Sample Preparation Kits v2 (Illumina, San Diego, Calif) and clustered onto a flowcell by using a cBOT amplification system with a HiSeq SR v4 Cluster Kit (Illumina). Single-read sequencing was carried out on a HiSeq2500 sequencer

(Illumina) by using a HiSeq SBS v4 Kit to generate 58-base reads, with a target of approximately 10 million reads per sample.

RNA sequencing data analysis

FASTQ files were downloaded from the website <https://basespace.illumina.com>. Libraries were processed via Galaxy^{E3} and aligned via TopHat (version 1.4.1)^{E4} to the human reference genome, Ensembl's Homo sapiens GRCh38, version 77 (GRCh38.77.gtf). The single-paired flag was set to single, whereas all other TopHat parameters were set to the default settings. HTSeq-count^{E5} was used to generate gene counts with the mode as intersection (nonempty) and minimum alignment quality set to 0 and otherwise set to the default parameters, resulting in 60,705 genes. The data set was subsequently set to include only the subset of protein-coding genes (N = 21,794) and then normalized via the median ratio normalization method in DeSeq2.^{E6} Finally, only genes with at least 1 count in 10% of the libraries were included in the analysis (N = 15,872).

Statistical analysis

All analyses were carried out in the language R.^{E7} Descriptive statistics were created to summarize relevant demographics of the cohort. Chi-square tests were used for categorical variables, and Kruskal-Wallis tests were used for continuous variables.

Our main aim was to evaluate classified gene sets over time for a phenotype based on asthma in children at 7 years of age. We utilized PLS-DA from the R package mixOmics,^{E8} a supervised method that is known for capturing the variability of both explanatory and dependent variables and reducing the relationship to a final optimized subset of genes that best differentiate the phenotype outcomes.^{E9} In this process we first assessed general performance of our gene set by using a regular PLS-DA analysis to get a sense of the initial model performance and input parameters for the subsequent sparse PLS-DA models. We tuned the sparse PLS-DA models and used a 5-fold cross-validation repeated 50 times for minimal stability and to thus ensure a robust model, as suggested in the mixOmics package.^{E8} We assessed the optimal number of components by looking at the classification error rates by component. Because variability in error rates was not high between the first 3 components, a decision was made to keep 3 components per analysis to assess the gene sets that were most predictive of 7-year phenotype. We then evaluated the performance of our models by again using 5-fold cross-validation repeated 50 times. Finally, we looked at the top 15 genes selected for a given component, which were quantitatively determined to show the most relative importance in a given component and specific phenotype. We fit 4 separate sparse PLS-DA models by using the 15,872 normalized gene set for the 60 subjects who had RNA sequencing data at all 3 years. As we observed high variability in individual gene expression between years 2, 5, and 7, we first created an aggregate view of all time points by calculating the geometric mean by subject and gene across all 3 time points, with use of the 7-year asthma phenotype variables as dependent outcome. Second, we fit a sparse PLS-DA model on the 15,872 normalized gene set for each year separately (ie, years 2, 5, and 7) for the 60 subjects who had a measurement at all 3 years.

REFERENCES

- E1. Gern JE, Visness CM, Gergen PJ, Wood RA, Bloomberg GR, O'Connor GT, et al. The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. *BMC pulmonary medicine* 2009;9:17.
- E2. O'Connor GT, Lynch SV, Bloomberg GR, Kattan M, Wood RA, Gergen PJ, et al. Early-life home environment and risk of asthma among inner-city children. *J Allergy Clin Immunol* 2018;141:1468-75.
- E3. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* 2016;44:W3-10.
- E4. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 2012;7:562-78.
- E5. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166-9.

- E6. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
- E7. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.
- E8. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* 2017;13:e1005752.
- E9. Le Cao KA, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* 2011;12:253.