Gene expression profiling of asthma phenotypes demonstrates molecular signatures of atopy and asthma control

Judie A. Howrylak, MD, PhD,[a] Matthew Moll, MD,[b] Scott T. Weiss, MD,[c,d,e] Benjamin A. Raby, MD,[c,d,e] Wei Wu, PhD,[f] and Eric P. Xing, PhD[g] *Hershey and Pittsburgh, Pa, and Boston, Mass*

Background: Recent studies have used cluster analysis to identify phenotypic clusters of asthma with differences in clinical traits, as well as differences in response to therapy with anti-inflammatory medications. However, the correspondence between different phenotypic clusters and differences in the underlying molecular mechanisms of asthma pathogenesis remains unclear.

Objective: We sought to determine whether clinical differences among children with asthma in different phenotypic clusters corresponded to differences in levels of gene expression.

Methods: We explored differences in gene expression profiles of $CD4^+$ lymphocytes isolated from the peripheral blood of 299 young adult participants in the Childhood Asthma Management Program study. We obtained gene expression profiles from study subjects between 9 and 14 years of age after they participated in a randomized, controlled longitudinal study examining the effects of inhaled anti-inflammatory medications over a 48-month study period, and we evaluated the correspondence between our earlier phenotypic cluster analysis and subsequent follow-up clinical and molecular profiles.

Results: We found that differences in clinical characteristics observed between subjects assigned to different phenotypic clusters persisted into young adulthood and that these clinical differences were associated with differences in gene expression patterns between subjects in different clusters. We identified a subset of genes associated with atopic status, validated the presence of an atopic signature among these genes in an independent cohort of asthmatic subjects, and identified the presence of common transcription factor binding sites corresponding to glucocorticoid receptor binding.

Conclusion: These findings suggest that phenotypic clusters are associated with differences in the underlying pathobiology of asthma. Further experiments are necessary to confirm these findings. (J Allergy Clin Immunol 2016;■■■:■■■-■■■.)

Key words: *Childhood asthma, asthma phenotypes, gene expression profiling, microarray analysis, longitudinal study*

Asthma is a disease of increased airway hyperresponsiveness and airflow limitation that is increasingly being viewed as a heterogeneous syndrome composed of an assortment of disease subtypes with differing causes and natural histories.[1] The observation that subsets of asthmatic patients exist who continue to have symptoms despite maximal medical therapy has motivated the search for distinct asthma subgroups with putative differences in disease mechanisms. Recent multivariate analyses have uncovered phenotypic clusters with differing risk factors for and manifestations of asthma. Moore et al[2] demonstrated the presence of 5 distinct phenotypic clusters among adult asthmatic patients, and Fitzpatrick et al[3] performed a parallel analysis with analogous findings among patients with childhood asthma. More recent work has highlighted the clinical importance of such clusters by demonstrating the presence of both longitudinal consistency[4] and different responses to medical therapy[5] between different phenotypic clusters.

An important implication of recent advances in our understanding of asthma phenotypes is that we can use these clusters to uncover associated differences in pathogenetic mechanisms and thus have the potential to identify new therapeutic targets with increased treatment specificity and new molecular biomarkers for improved clinical detection.

Several studies have furthered our current understanding of the relationship between phenotypic clusters and molecular mechanism. Woodruff et al[6] profiled a selected subset of gene expression levels in asthmatic patients and found that differences in gene expression corresponded to differences in multiple clinical measures of asthma severity, demonstrating a link between clinical phenotype and molecular mechanism. Baines et al[7] subsequently found a correspondence between transcriptional profiles and different clinical characteristics in an asthmatic population. However, the cross-sectional nature of these studies limits the clinical applicability of the findings.

In the current analysis our goal was to link differences in gene expression levels to longitudinally stable clinical phenotypes with demonstrated differences in response to medical therapy. In a prior analysis we determined the presence of phenotypic clusters in a cohort of children with mild-to-moderate persistent asthma

From [a]the Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, Penn State Milton S. Hershey Medical Center, Hershey; [b]the Department of Medicine, Boston University; [c]Harvard Medical School, Boston; [d]the Channing Division of Network Medicine and [e]the Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston; and [f]the Computational Biology Department and [g]the Department of Machine Learning, School of Computer Science, Carnegie Mellon University, Pittsburgh.

Supported by grant R01 HL086601 from the National Heart, Lung, and Blood Institute/National Institutes of Health (NIH/NHLBI). The Childhood Asthma Management Program Genetics Ancillary Study is supported by U01 HL075419, U01 HL65899, P01 HL083069, and T32 HL07427 from the NIH/and NHLBI. This article is subject to the National Institutes of Health Public Access Policy (http://publicaccess.nih.gov).

Disclosure of potential conflict of interest: W. Wu has received research support from the National Institutes of Health (NIH; grant R01 GM093156). E. P. Xing has received research support from the NIH (grant NIH GWAS R016M087694). The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication March 2, 2015; revised August 13, 2015; accepted for publication September 30, 2015.

Corresponding author: Judie A. Howrylak, MD, PhD, Department of Medicine, Penn State Hershey Medical Center, 500 University Dr, Hershey, PA 17033. E-mail: jhowrylak@hmc.psu.edu.

0091-6749/$36.00

http://dx.doi.org/10.1016/j.jaci.2015.09.058

Abbreviations used

ADD3:	Adducin 3 (gamma)
CAMP:	Childhood Asthma Management Program
DE:	Differentially expressed
FDR:	False discovery rate
GR:	Glucocorticoid receptor
HDAC2:	Histone deacetylase 2
SLC33A1:	Solute carrier family 33, member 1
SRM:	Spermidine synthase
TFBS:	Transcription factor binding site

obtained from the Childhood Asthma Management Program (CAMP) study.[8] Among these children, we identified 5 distinct phenotypic clusters with different degrees of airflow obstruction, rates of exacerbation, and atopic characteristics. We further found that these clusters demonstrated both longitudinal stability over the 48-month study period and differences in response to medical therapy. In the current study we extend our earlier analysis through an exploration of differences in gene expression between these different phenotypic clusters, with the goals of identifying novel molecular biomarkers corresponding to different phenotypes and further elucidating the differences in molecular mechanism between subjects in different clusters.[5] We uncovered the presence of a set of genes in $CD4^+$ lymphocytes isolated from the peripheral blood of a subset of CAMP participants who were differentially expressed (DE) between more atopic and less atopic study subjects. Gene expression levels were also associated with different phenotypic clusters and were highly predictive of multiple clinical characteristics, such as levels of atopy and asthma control. We validated these results in an independent population and evaluated for the presence of shared transcription factor binding sites (TFBSs) among the genes of each module.

METHODS

Study population

CAMP was a multicenter, randomized, double-masked clinical trial of the long-term effects of 3 inhaled treatments for mild-to-moderate childhood asthma with 1041 subjects enrolled.[9] Recruitment for the CAMP study took place from December 1993 to September 1995. Two subsequent 4-year observational follow-up studies of CAMP participants, CAMPCS/1 and CAMPCS/2, were carried out on completion of the original CAMP study. We obtained blood samples and clinical data for the current study during routine CAMPCS/2 clinic visits between May 1, 2004, and July 31, 2007, from 4 clinical centers (Baltimore, Boston, Denver, and St Louis). The study visit included questionnaire assessments of asthma symptoms and medication use. From those specimens, we isolated $CD4^+$ lymphocytes and obtained high-quality expression profiles from 299 patient samples using Illumina HumanRef8 v2 BeadChip arrays (Illumina, San Diego, Calif).[10] Clinical characteristics and gene expression profiles of the 299 study subjects were assessed with respect to their membership in one of 5 phenotypic clusters assigned in a prior analysis of the complete CAMP cohort of 1041 study subjects.[5]

RNA extraction and microarray preprocessing

From the CAMP study population, we isolated $CD4^+$ T cells from the collected mononuclear cell layer using anti-$CD4^+$ microbeads with column separation (Miltenyi Biotec, Auburn, Calif).[11,12] Total RNA was extracted with the RNeasy Mini Protocol (Qiagen, Gaithersburg, Md).[13-15] Expression profiles were generated with the Illumina Human-Ref8 v2 BeadChip arrays (Illumina) according to the protocol. Arrays were read with the Illumina BeadArray scanner and analyzed by using BeadStudio (version 3.1.7) without background correction. Raw expression intensities were processed with the *lumi* package[16] of Bioconductor, with background adjustment with Robust Multi-Array Average convolution[17] and $\log_2$ transformation of each array. The combined samples were quantile normalized. The complete raw and normalized microarray data are available through the Gene Expression Omnibus of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/, accession ID GSE22324).

Identification of DE genes

To classify gene expression levels from multiple phenotypic clusters into differential expression patterns, we used an empiric Bayes hierarchical modeling approach to calculate the posterior probability of each gene expression value fitting a particular pattern of expression.[18-20] For example, for this analysis, we were interested in patterns of differential expression of genes across different phenotypic clusters. We developed a set of 49 theoretical pattern assumptions (Fig 1 and see Table E1 in this article's Online Repository at www.jacionline.org), such as the assumption of the null hypothesis of no differential expression across clusters for a gene or the assumption of differential expression across all clusters for a gene, and then calculated the posterior probability of each gene fitting a particular pattern of expression. We assigned genes to the gene pattern with maximum posterior probability. Further details of this methodology are described in the Methods section in this article's Online Repository at www.jacionline.org.

Association of differential gene expression with clinical characteristics

To assess the potential functional relevance of the DE genes, we explored the relationship between the top pattern of differential expression and multiple clinical characteristics obtained at the time blood samples were obtained for gene expression profiling. We used Kruskal-Wallis and χ^2 tests to make comparisons between phenotypic clusters with different gene expression profiles. We calculated counts and percentages or arithmetic means and SDs for all variables measured. We also examined the temporal effects of differential gene expression by calculating correlations between gene expression levels and several longitudinal clinical outcomes, including measures of atopy and airway hyperresponsiveness.

Validation of gene expression signatures

To assess the generalizability of the association between atopy and gene expression levels, we evaluated whether the genes fitting the atopic expression pattern could be used to predict atopic status in an independent cohort. We used a gene expression data set that was publicly available on the Gene Expression Omnibus Web site (accession no. GSE473).[21] This data set consisted of gene expression profiles obtained from $CD4^+$ T lymphocytes in the peripheral blood of 30 patients with and without atopy and asthma. We selected this data set as a validation cohort because it was similar to our study population in terms of the range of clinical phenotypes (atopic and asthmatic subjects) and in terms of the particular cell type from which RNA was obtained ($CD4^+$ T lymphocytes). Notably, this cohort was somewhat different from our study population because only a subset of patients had asthma (68/88), whereas all of the 299 subjects in our study population had asthma.

We used the genes DE between more atopic and less atopic clusters to grow a binary recursive partitioning decision tree to predict phenotypic cluster assignments within our patient population.[22,23] Further details of this methodology are described in the Methods section in this article's Online Repository.

Identification of common regulatory domains

To map the set of 501 DE genes to biological functions, we used the ENCODE ChIP-Seq Significance Tool[24] to explore the presence of enriched transcription factors within each module. The ENCODE ChIP-Seq

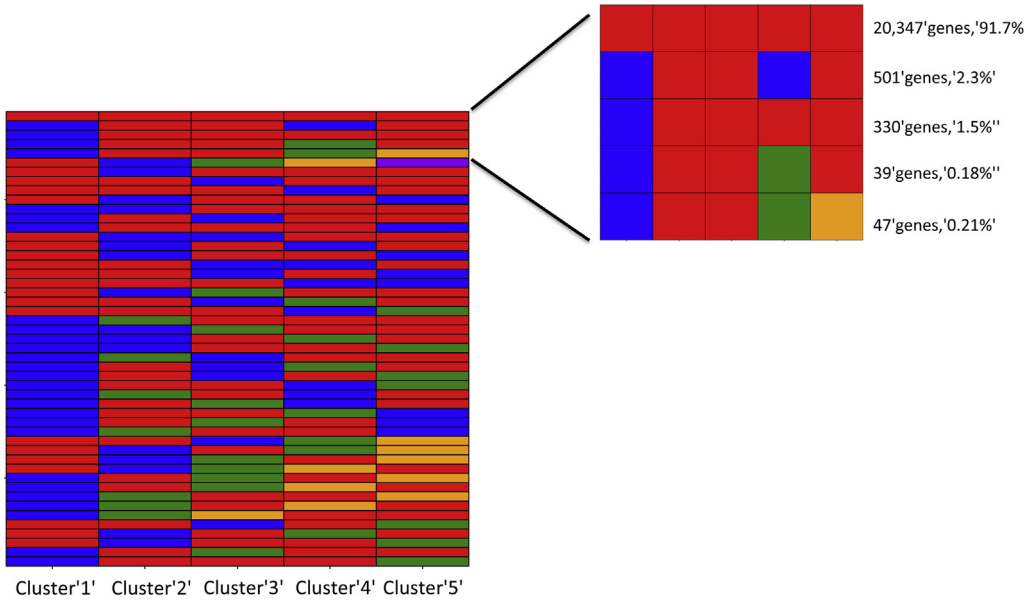


FIG 1. Graphic depiction of possible distributions of gene expression patterns for each phenotypic cluster. Each *color* denotes a particular genome-wide pattern of expression, and each *row* denotes a distribution of expression patterns across phenotypic clusters. Similar colors denote similar patterns of expression, whereas different colors denote different patterns of expression. For example, the top row of the figure depicts 5 red segments, indicating similar patterns of genome-wide expression across the 5 phenotypic clusters. A close-up view of the top 5 expression pattern distributions as determined by posterior probability is also shown.

Significance Tool leverages a MySQL database of official unified peak calls from 708 ENCODE ChIP-Seq nonhistone and noncontrol experiments, including 220 transcription factor and treatment combinations across 91 cell types. We designed the experiment so that for each transcription factor/treatment combination, the ENCODE ChIP-Seq Significance Tool would query the database to identify the number of genes within each coexpression module with at least 1 transcription factor peak apex in the selected window around the gene. We calculated an enrichment score by using a 1-tailed hypergeometric test followed by the false discovery rate (FDR) method to control the FDR in multiple hypothesis testing.[25]

RESULTS

Distribution of phenotypic traits

Clinical phenotypic data were available for all 299 participants. We assessed clinical characteristics of study participants based on their membership in a particular phenotypic cluster, as previously assigned. The characteristics assessed within 1 month of the time blood was obtained for microarray analysis are presented in Table I. The clinical characteristics presented in Table I represent follow-up data obtained between 9 and 14 years after the onset of the original CAMP study. The mean age of study subjects was 20.4 years compared with 5 to 12 years in the original study. The ethnic and sex distributions were similar to those of the original study (see Table E2 in this article's Online Repository at www.jacionline.org).

Several measures of atopic burden were obtained at the time of sample collection, including serum IgE levels and eosinophil counts. As was observed in our prior cluster analysis, the degree of atopy at the time of follow-up was highest in clusters 2, 3, and 5 and lowest in clusters 1 and 4. Similarly, the spirometric values at the time of follow-up show a similar distribution to those obtained at baseline, with clusters 4 and 5 showing the highest levels of airway obstruction.

The number of active smokers represented a minority of this cohort (11.4%), with the highest percentage of smokers in cluster 3 (16.3%) and the lowest percentage of smokers in cluster 4 (4.3%).

Gene transcripts demonstrate patterns of expression associated with atopic status

To understand the relative contribution of different genes to the formation of the asthma phenotypic clusters, we performed gene expression profiling of subjects from different phenotypic clusters to detect patterns of expression. For the set of phenotypic clusters, gene expression levels could be sorted into 49 distinct theoretical patterns (Fig 1 and see Table E1).

For each transcript in each phenotypic cluster, we calculated the posterior probability for each of the 49 patterns and assigned the transcript to the expression pattern with maximum posterior probability. DE transcripts were defined as those with maximum posterior probabilities of greater than a specific threshold set to limit the FDR to less than 0.05 for each of the DE patterns (Fig 1, and see pattern numbers 2-49 in Table E1). Using this approach, we found that 99.7% of the DE transcripts were confined to 2 of the 49 possible DE patterns.

The top 5 expression patterns as ranked by posterior probability are also shown in Fig 1. The expression pattern containing the highest number of DE transcripts (20,347 of 22,184 total transcripts) was the null hypothesis expression pattern (ie, the pattern of no difference in expression between the different phenotypic clusters). The expression pattern containing the second highest number of DE transcripts (501 of 22,184 total

TABLE I. Distribution of traits across cluster groups

	Cluster 1 (n = 102)	Cluster 2 (n = 50)	Cluster 3 (n = 49)	Cluster 4 (n = 70)	Cluster 5 (n = 28)	
AOE classification:	**LLL**	**HLL**	**HHM**	**MHH**	**HHH**	***P* value**
Demographics						
Age (y)	20.1 ± 2.0	20.3 ± 2.2	21.0 ± 2.2	20.7 ± 2.1	19.9 ± 2.4	.54
Male (%)	59 (57.8)	37 (74.0)	24 (49.0)	44 (62.9)	20 (71.4)	.08
Female (%)	43 (42.2)	13 (26.0)	25 (51.0)	26 (37.1)	8 (28.6)	
White (%)	84 (82.4)	35 (70.0)	34 (69.4)	53 (75.7)	21 (75.0)	.16
African American (%)	11 (10.8)	13 (26.0)	10 (20.4)	16 (22.9)	6 (21.4)	
Hispanic (%)	7 (6.9)	2 (4.0)	5 (10.2)	1 (1.4)	1 (3.6)	
Atopic features						
Total serum IgE levels ($\log_{10}$)	0.63	0.70	0.64	0.80	0.66	.001
Total serum eosinophils ($\log_{10}$)	2.17 ± 0.50	2.37 ± 0.46	2.42 ± 0.35	2.34 ± 0.32	2.44 ± 0.31	.002
Spirometry						
Prebronchodilator FEV_1 (% predicted)	98.2 ± 13.2	98.5 ± 10.9	98.4 ± 11.3	94.1 ± 13.7	97.0 ± 11.1	.48
Prebronchodilator FEV_1/FVC ratio	78.5 ± 7.67	78.5 ± 7.89	78.6 ± 7.55	76.2 ± 7.57	75.3 ± 9.64	.22
Prebronchodilator peak flow (L/min)	576.6 ± 144.5	634.0 ± 157.1	564.0 ± 147.0	579.8 ± 137.8	598.8 ± 162.7	.20
Airway responsiveness						
Methacholine PC_{20} (natural log)	1.09 ± 0.56	0.91 ± 0.53	0.75 ± 0.47	0.92 ± 0.52	0.96 ± 0.58	.08
Environmental exposures						
Tobacco smoking						.18
Yes	14 (13.7)	7 (14.0)	8 (16.3)	3 (4.3)	2 (7.1)	
No	76 (74.5)	38 (76.0)	34 (69.4)	60 (85.7)	21 (75.0)	
Average cigarettes smoked per day	1.3 ± 4.1	1.2 ± 3.7	1.9 ± 4.7	0.5 ± 2.8	0.5 ± 2.1	.18

We constructed an Atopy-Obstruction-Exacerbation (AOE) scoring scheme, assigning low (L), medium (M), or high (H) scores for each factor group. *FVC*, Forced vital capacity.

transcripts) was the pattern of similar expression between clusters 1 and 4 and between clusters 2, 3, and 5. Our earlier analysis of the clinical data from the CAMP study demonstrated that clusters 1 and 4 had the lowest atopic burden of the phenotypic clusters. That is, at the time of initial recruitment to the CAMP study, clusters 1 and 4 had the lowest levels of atopic dermatitis (cluster 1 and cluster 4 = 0%), the lowest history of hay fever (cluster 1 = 20.3%, cluster 4 = 52.9%), the lowest history of a positive skin test result (cluster 1 = 76.7%, cluster 4 = 88%), and the lowest $\log_{10}$ total serum IgE levels (cluster 1 = 2.37, cluster 4 = 2.64). Thus a large number of DE genes demonstrated an expression pattern that was associated with the degree of atopic burden present among study subjects at the time of enrollment in the CAMP study, suggesting atopic status was the primary driver of the change in gene expression for these transcripts.

Gene expression profiles are associated with differences between clinical traits

We explored the relationship between gene expression pattern and clinical traits by reorganizing our 5 phenotypic clusters into 2 subgroups (most atopic and least atopic) based on the degree of atopic burden present. These 2 subgroups were associated with the pattern of expression of 501 genes that differed between subjects with different levels of atopy. We examined multiple clinical characteristics for differences between the 2 subgroups and found statistically significant differences in serum IgE levels ($P = .005$), serum eosinophil counts ($P = .0002$), and established markers for atopic burden (Table II). We also found statistically significant differences in methacholine PC_{20}, a measure of airway hyperresponsiveness between the 2 subgroups. The group with the higher burden of atopy demonstrated a lower methacholine PC_{20} value ($P = .04$), indicating that in addition to having higher levels of atopy as adults, this subgroup also had higher levels of airway hyperresponsiveness.

TABLE II. Characteristics of study subjects dichotomized by atopic status

	More atopic (n = 127)	Less atopic (n = 172)	*P* value
Demographics			
Age (y)	20.5 ± 2.3	20.4 ± 2.0	.63
Male (%)	81 (63.8)	103 (59.9)	.55
Female (%)	46 (36.2)	69 (40.1)	
White (%)	90 (70.9)	137 (80.0)	.21
African American (%)	29 (22.8)	27 (15.7)	
Hispanic (%)	8 (6.3)	8 (4.7)	
Atopic features			
Serum IgE ($\log_{10}$)	2.69 ± 0.57	2.43 ± 0.63	.005
Serum eosinophils ($\log_{10}$)	2.41 ± 0.38	2.24 ± 0.44	.0002
Spirometry			
Prebronchodilator FEV_1 (% predicted)	98.1 ± 11.0	96.5 ± 13.5	.29
Postbronchodilator FEV_1/FVC ratio (% predicted)	77.8 ± 8.22	77.4 ± 7.69	.82
Airway responsiveness			
Methacholine PC_{20} (natural log)	0.86 ± 0.52	1.03 ± 0.55	.04
Environmental exposures			
Tobacco smoking (%)			.40
Yes	17 (13.4)	17 (10.0)	
No	93 (73.2)	136 (79.1)	

FVC, Forced vital capacity.

We next examined the longitudinal trajectory of each clinical trait comparing more atopic and less atopic subjects. In general, for each subgroup, levels of atopic burden decreased over time, as exhibited by decreasing serum IgE levels and serum eosinophil counts (Fig 2). In addition, levels of airway hyperresponsiveness decreased over time, as demonstrated by increasing methacholine PC_{20} values. However, we also found that along the course of each trait, there were differences between the more atopic and less

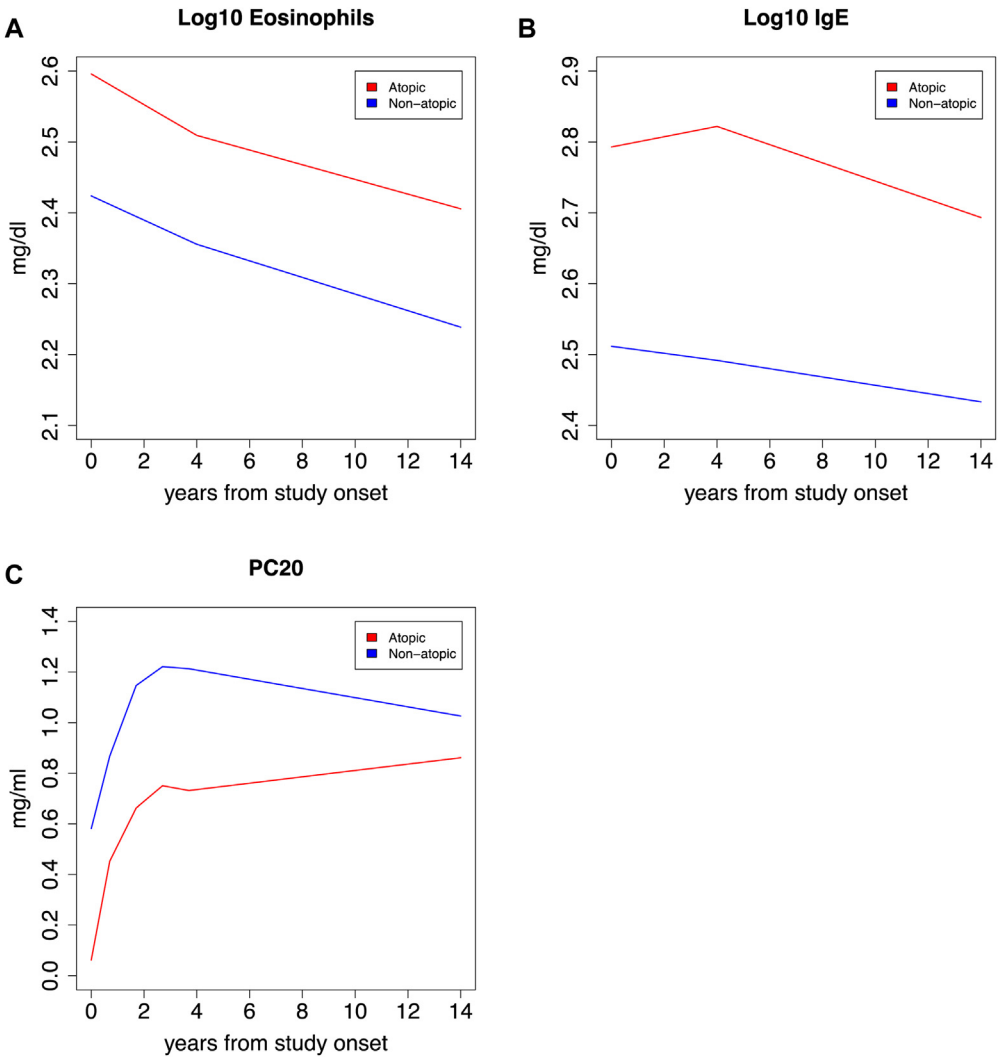


FIG 2. Longitudinal trajectories of several clinical traits stratified by gene expression pattern (more atopic vs less atopic). **A,** Trajectory of serum eosinophil counts. **B,** Trajectory of serum IgE levels. **C,** Trajectory of methacholine PC_{20} values.

atopic subgroups ($P < .001$). Overall, the differences between the more atopic and less atopic subgroups were more extreme in childhood, decreasing over time yet persisting into young adulthood.

Gene expression profiles are predictive of atopic status

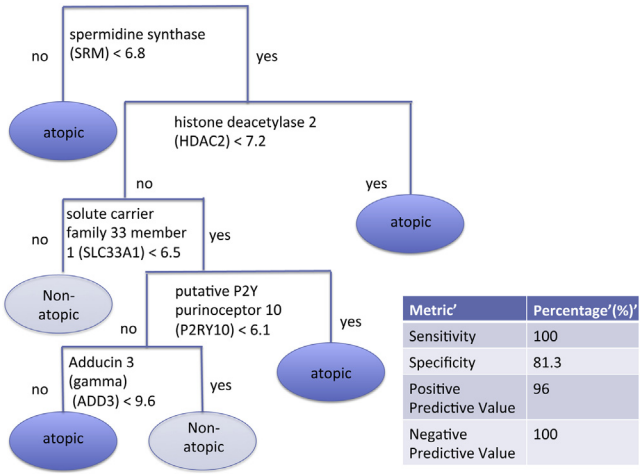
To explore the generalizability of the set of DE genes within our cohort, we assessed whether these genes were also predictive of atopic status in an independent data set (Fig 3). Of the 501 genes present in this gene expression pattern, 5 genes successfully classified all of the atopic patients in an independent cohort with a sensitivity of 100%, specificity of 81.3%, positive predictive value of 96.0%, and negative predictive value of 100%. Details of the methodology leading to these results are provided in the Methods section in this article's Online Repository.

Expression levels for each of the 5 genes in the signature for atopic status were significantly different ($P < .05$) between more atopic and less atopic subgroups (Table III). Spermidine synthase *(SRM)* expression levels were relatively downregulated in more atopic compared with less atopic asthmatic patients, whereas expression levels of histone deacetylase 2 *(HDAC2)*; solute carrier family 33, member 1 *(SLC33A1)*; purinergic receptor P2Y, G-protein coupled, 10 *(P2RY10)*; and adducin 3 (gamma *[ADD3]*) were relatively upregulated in more atopic compared with less atopic asthmatic patients.

We measured the correlation between the 5-gene signature for atopic status and serum IgE levels, serum eosinophil counts, and methacholine PC_{20} values. Serum IgE levels were positively correlated with all 5 genes in the gene signature, most notably with *ADD3* ($P = .04$, Fig 4), and serum eosinophil counts were positively correlated with *SRM* expression levels ($P = .02$).

DE genes are enriched for glucocorticoid receptor binding sites

We queried the ENCODE Chip-Seq Significance Tool for evidence of TFBS enrichment among our set of 501 DE genes. Among the genes expressing atopic pattern differences, we found



Metric'	Percentage'(%)'
Sensitivity	100
Specificity	81.3
Positive Predictive Value	96
Negative Predictive Value	100

FIG 3. Decision tree classification model for atopic status. There were 5 genes that were strongly predictive of atopic status in an independent population. These genes make up the branches of the tree, with each partition determined by the $\log_2$ gene expression level.

TABLE III. Average expression levels of atopic signature genes

Gene	More atopic (n = 127)	Less atopic (n = 172)	*P* value	Adjusted *P* value
SRM	9.11 ± 0.53	9.27 ± 0.57	.019	.019
HDAC2	10.5 ± 0.51	10.3 ± 0.63	.0048	.0060
Solute carrier family 33, member 1 *(SLC33A1)*	8.4 ± 0.40	8.2 ± 0.42	.0018	.0060
Purinergic receptor P2Y, G-protein coupled, 10 *(P2RY10)*	9.4 ± 0.71	9.2 ± 0.70	.0027	.0060
ADD3	11.3 ± 0.91	11.0 ± 0.93	.0048	.0060

evidence that the promoter regions were significantly enriched for glucocorticoid receptor (GR) binding sites. GR binding sites were present in 144 of 476 genes queried by using the tool (q value = 4.99e-9). The list of genes with GR binding sites is shown in Table E3 in this article's Online Repository at www.jacionline.org. Three of the 5 genes present in the signature (*SRM*, *SLC33A1*, and *ADD3*) were among those with TFBSs enriched for GR binding sites.

DISCUSSION

Our analysis of gene expression profiles obtained from CAMP participants 9 to 14 years of age from study onset is notable for several key findings. First and foremost, these results extend and enhance the results of our previous cluster analysis. In the previous cluster analysis we detected 5 phenotypic clusters using the baseline clinical data from asthmatic study subjects who differed in their response to medical therapy. In the current analysis we found that in the same patient population, even after 9 to 14 years, there continued to be longitudinal consistency in the clinical characteristics of subjects within different phenotypic clusters.

Second, subjects with a higher degree of atopic features demonstrated differential expression in a subset of genes with expression levels highly correlated with longitudinal measures of atopy and airway hyperresponsiveness.

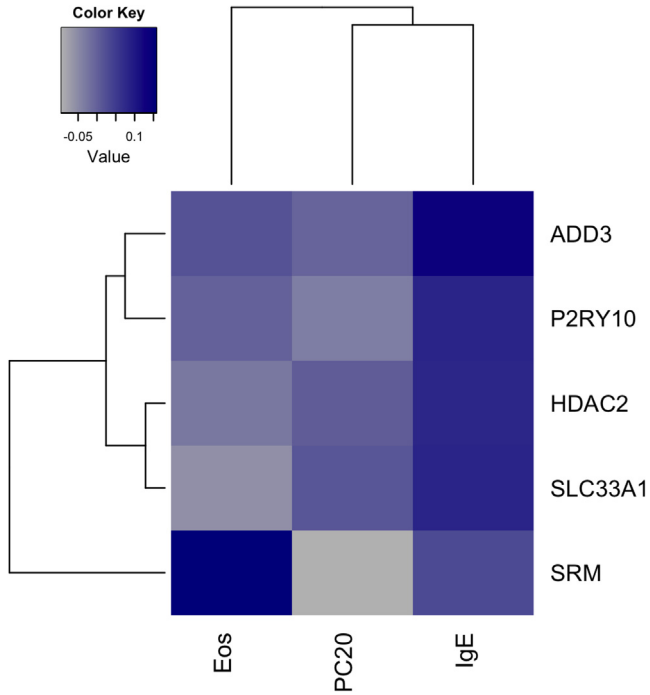


FIG 4. Correlation heat map depicting correlation between clinical traits and gene expression levels. Blue denotes a higher positive level of correlation between gene and clinical trait, whereas grey denotes a higher negative level of correlation between gene and clinical trait.

Third, a subset of these DE genes formed an atopic signature that we used to successfully determine atopic status from gene expression profiles obtained from an independent population of asthmatic patients.

In our initial cluster analysis performed with clinical data from participants in the CAMP study, we found that children could be characterized in terms of 5 distinct phenotypic clusters, which differed in terms of atopic burden, airway obstruction, and rates of exacerbation. In this analysis we found that an average of 12 years after the original study, participants continued to exhibit cluster-specific differences in several clinical characteristics, including differences in atopic features, spirometry, and airway responsiveness. This is notable because, at this time, in contrast to the physiologic differences, many of these patients described relatively mild symptoms. Furthermore, this finding, from the largest randomized, placebo-controlled clinical trial with extended follow-up for children with mild-to-moderate asthma,[26] suggests that the decreased symptoms many patients with childhood asthma describe as they age do not correspond to disease remission. In fact, the original pathogenetic mechanisms appear to persist into young adulthood, although with minimal subjective symptoms.

Gene expression profiles from $CD4^+$ T cells collected from the same set of patients allow us to integrate genomic factors into our analysis of phenotypic clusters. Our analysis of gene expression profiles further confirms the persistence of physiologic differences between asthmatic patients assigned to different clusters. A set of 501 of 22,184 total genes assayed displayed an expression pattern that was associated with atopic status. Thus these gene expression profiles allowed us to organize our phenotypic clusters into 2 subgroups based on atopic burden. Study subjects with higher levels of atopic burden had a different pattern of expression

of this set of genes from subjects with relatively lower levels of atopic burden. Atopic subjects (clusters 2, 3, and 5) had higher levels of 2 peripheral blood markers of atopic burden, serum IgE levels and serum eosinophil counts, compared with the other clusters (clusters 1 and 4). Additionally, methacholine PC_{20} values were lower in more atopic than less atopic subjects, suggesting that the set of genes responsible for atopy might also control aspects of airway hyperresponsiveness. Differences in these clinical traits were significant when these children were initially enrolled in the CAMP study and persisted for at least 9 to 14 years into young adulthood.

A subset of DE genes formed a signature that was highly predictive of atopy in an independent population of patients with mild-to-moderate asthma, a finding that lends further credibility to the hypothesis that different phenotypic clusters correspond to differences in the underlying pathobiological mechanisms of asthma, as well as validating the biological relevance of our longitudinal phenotypic clusters. Several genes in this signature have been previously linked to asthma pathogenesis, most notably *HDAC2*, which had higher expression in more atopic relative to less atopic study subjects. Studies have shown that reduction of histone acetylation by *HDAC2* results in suppression of multiple inflammatory genes through a glucocorticoid-mediated process and a reduction in asthmatic symptoms.[27] Evidence has also linked inhibition of *HDAC2* to steroid-resistant asthma.[28-32] The presence of GR TFBSs on numerous DE genes suggests an association between steroid resistance and differences in the clinical characteristics we observed between our more atopic and less atopic subgroups.

Our study had several limitations. First, the analysis was retrospective, with gene expression profiles associated with existing clinical phenotypic data. The availability of longitudinal gene expression data from prospective studies could help to elucidate temporal changes in the expression of these genes and their role in the pathogenesis of childhood asthma.

Second is the lack of longitudinal clinical data in our validation cohort. Further studies from multiple institutions will be necessary to confirm the correspondence between changes in gene expression, atopic status, and clinical characteristics corresponding to atopy and airway hyperresponsiveness.

Third is the fact that our patient cohort was limited to subjects with mild and moderate asthma, excluding those with a more severe phenotype. We recognize that this limited the diversity of symptoms among our subjects, and many young adults among the more mild phenotypes seemed to outgrow their asthma symptoms. For this reason, the differential gene expression levels among different clusters are more likely to have clinical utility for patients at the extreme ends of the phenotype.

In summary, the current study represents the largest analysis to demonstrate differences in gene expression profiles associated with clinically significant phenotypic clusters. Furthermore, differences among patients in different clusters and their gene expression profiles persist across time. Our findings lend further support to the hypothesis that asthma phenotypic clusters are associated with differences in the underlying molecular mechanisms of asthma pathogenesis, as demonstrated by differences in gene expression profiles among subjects in different phenotypic clusters. This finding has implications for drug development and personalized approaches to the treatment of this complex disease. Further work will be necessary to validate these early findings and explore the mechanistic differences between different clusters.

We thank all subjects for their ongoing participation in this study. We acknowledge the CAMP investigators and research team, who were supported by the National Heart, Lung, and Blood Institute, for collection of CAMP Genetic Ancillary Study data. Special thanks go to Anne Plunkett, Teresa Concordia, Debbie Bull, Denise Rodgers, and D. Sundstrom for their assistance with sample collection. All work on data collected from the CAMP Genetic Ancillary Study was conducted at the Channing Laboratory of the Brigham and Women's Hospital and Harvard Medical School under appropriate CAMP policies and human subject's protections.

Clinical implications: There is evidence for asthma endotypes (subtypes) but less information about the relevance of such endotypes. In this article we link asthma endotypes to atopic patterns of gene expression.

REFERENCES

1. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. Nat Med 2012;18:716-25.
2. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med 2010;181:315-23.
3. Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, et al. Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. J Allergy Clin Immunol 2011;127:382-9.
4. Boudier A, Curjuric I, Basagana X, Hazgui H, Anto JM, Bousquet J, et al. Ten- year follow-up of cluster-based asthma phenotypes in adults. A pooled analysis of three cohorts. Am J Respir Crit Care Med 2013;188:550-60.
5. Howrylak JA, Fuhlbrigge AL, Strunk RC, Zeiger RS, Weiss ST, Raby BA. Clas- sification of childhood asthma phenotypes and clinical responses to inhaled anti-inflammatory medications. J Allergy Clin Immunol 2014;133:1289-300.
6. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, et al. T-helper type 2-driven inflammation defines major subphenotypes of asthma. Am J Respir Crit Care Med 2009;180:388-95.
7. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. J Allergy Clin Immunol 2011;127:153-60.
8. Long-term effects of budesonide or nedocromil in children with asthma. The Childhood Asthma Management Program Research Group. N Engl J Med 2000;343:1054-63.
9. The Childhood Asthma Management Program (CAMP): design, rationale, and methods. Childhood Asthma Management Program Research Group. Control Clin Trials 1999;20:91-120.
10. Murphy A, Chu JH, Xu M, Carey VJ, Lazarus R, Liu A, et al. Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. Hum Mol Genet 2010;19:4745-57.
11. Jonuleit H, Schmitt E, Schuler G, Knop J, Enk AH. Induction of interleukin 10-producing, nonproliferating CD4(+) T cells with regulatory properties by repetitive stimulation with allogeneic immature human dendritic cells. J Exp Med 2000;192:1213-22.
12. Zorn E, Miklos DB, Floyd BH, Mattes-Ritz A, Guo L, Soiffer RJ, et al. Minor histocompatibility antigen DBY elicits a coordinated B and T cell response after allogeneic stem cell transplantation. J Exp Med 2004;199:1133-42.
13. Chambers J, Angulo A, Amaratunga D, Guo H, Jiang Y, Wan JS, et al. DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. J Virol 1999;73:5757-66.
14. Gonzalez P, Zigler JS Jr, Epstein DL, Borras T. Identification and isolation of differentially expressed genes from very small tissue samples. Biotechniques 1999;26:884-6, 888-92.
15. Gu L, Tseng S, Horner RM, Tam C, Loda M, Rollins BJ. Control of TH2 polarization by the chemokine monocyte chemoattractant protein-1. Nature 2000;404:407-11.
16. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547-8.
17. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high-density oligonucleotide array probe level data. Biostatistics 2003;4:249-64.

18. Newton MA, Kendziorski CM, Richmond CS, Blattner FR, Tsui KW. On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. J Comput Biol 2001;8:37-52.
19. Kendziorski CM, Chen M, Yuan M, Lan H, Attie AD. Statistical methods for expression quantitative trait loci (eQTL) mapping. Biometrics 2006;62:19-27.
20. Yuan M, Kendziorski C. A unified approach for simultaneous gene clustering and differential expression identification. Biometrics 2006;62:1089-98.
21. Mobini R, Andersson BA, Erjefalt J, Hahn-Zoric M, Langston MA, Perkins AD, et al. A module-based analytical strategy to identify novel disease-associated genes shows an inhibitory role for interleukin 7 Receptor in allergic inflammation. BMC Syst Biol 2009;3:19.
22. Breiman L. Classification and regression trees. Belmont (CA): Wadsworth International Group; 1984.
23. Ripley BD. Pattern recognition and neural networks. Cambridge: Cambridge University Press; 1996.
24. Auerbach RK, Chen B, Butte AJ. Relating genes to function: identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool. Bioinformatics 2013;29:1922-4.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Series B Stat Methodol 1995;57:289-300.
26. Covar RA, Fuhlbrigge AL, Williams P, Kelly HW. Childhood Asthma Management Program Research Group. The Childhood Asthma Management Program (CAMP): contributions to the understanding of therapy and the natural history of childhood asthma. Curr Respir Care Rep 2012;1:243-50.
27. Barnes PJ. Corticosteroid resistance in patients with asthma and chronic obstructive pulmonary disease. J Allergy Clin Immunol 2013;131:636-45.
28. Li LB, Leung DYM, Martin RJ, Goleva E. Inhibition of histone deacetylase 2 expression by elevated glucocorticoid receptor beta in steroid-resistant asthma. Am J Respir Crit Care Med 2010;182:877-83.
29. Ito K, Chung KF, Adcock IM. Update on glucocorticoid action and resistance. J Allergy Clin Immunol 2006;117:522-43.
30. Adcock IM, Ford PA, Bhavsar P, Ahmad T, Chung KF. Steroid resistance in asthma: mechanisms and treatment options. Curr Allergy Asthma Rep 2008;8:171-8.
31. Barnes PJ, Adcock IM. Glucocorticoid resistance in inflammatory diseases. Lancet 2009;373:1905-17.
32. Barnes PJ. Mechanisms and resistance in glucocorticoid control of inflammation. J Steroid Biochem Mol Biol 2010;120:76-85.

METHODS

Identification of DE genes

To identify distinct patterns within our set of gene expression profiles, we first postulated different potential patterns of gene expression based on variable patterns of gene expression among our phenotypic clusters. For example, for one pattern of expression, we postulated that gene expression across all phenotypic clusters would be similar. For another pattern of expression, we postulated that gene expression would be different for each cluster. For a third pattern of expression, we postulated that gene expression would be different among atopic and nonatopic asthmatic subjects. In total, we postulated 49 unique patterns of differential gene expression (Fig 1 and Table E1). Next, we used the previously described method[E1-E6] to calculate the posterior probability for each gene fitting a particular pattern of expression.

We used the EBarrays package in Bioconductor, which used a parametric empiric Bayes method to identify DE genes. For example, suppose that we want to compare expression levels between 2 prespecified expression patterns, s_1 and s_2, assumed by expression measurements $\mathrm{x}j = (x_{j1}, x_{j2},\ldots, x_{jI})$ taken on gene j. In the case where the distribution of measured expression levels is not affected by the groupings s_1 and s_2, we might conclude that the null hypothesis holds and that there is equivalent expression, $\mathrm{EE_j}$, for gene j. Conversely, if differential expression levels (DEj) exist, we might assume 2 different means, μ_{j1} and μ_{j2}, corresponding to measurements in s_1 and s_2, respectively. An additional assumption of this particular model is that the effects of each gene arise from independently and identically from a system-specific distribution, $\pi(\mu)$, which allows for information sharing among different genes.

In this model, if we let p signify the subset of genes that are DE, then $1 - p$ is the fraction of genes that are equivalently expressed (EE), such that the EE gene j presents data $\mathbf{x}j = (x_{j1}, x_{j2},\ldots, x_{jI})$, according to a distribution as follows:

$$\mathrm{f}_0(\mathbf{x}_j) = \int \left(\prod_{i=1}^{I} f_{obs}(x_{ji}) | \mu \right) \pi(\mu) d\mu.$$

Conversely, if gene j is DE, the data $\mathbf{x}j = (x_{j1}, x_{j2},\ldots, x_{jI})$ follow the following distribution:

$$f_1(\mathbf{x}_j) = f_0(\mathbf{x}_{j1}) f_0(\mathbf{x}_{j2}).$$

which demonstrates that x_{j1} and x_{j2} have different mean values. By extension, the marginal distribution of the data is as follows:

$$pf_1(\mathbf{x}_j) + (1-p) f_0(\mathbf{x}_j).$$

The posterior probability of differential expression as calculated by Bayes' rule and given the estimates of p, f_0, and f_1 is as follows:

$$\frac{pf_1(\mathbf{x}_j)}{pf_1(\mathbf{x}_j) + (1-p) f_0(\mathbf{x}_j)}.$$

In practice each gene is assigned to the distribution for which the posterior probability is greatest.

Validation of gene expression signatures

We evaluated the predictive potential of the set of 501 genes that were DE between atopic and nonatopic study subjects. We used these genes as predictors of atopic status in a publicly available independent cohort of patients with and without atopy and asthma. Using the methods described above, we identified a set of 501 genes that were highly associated with atopic status in our population. Next, we used the 501 genes from our data set to 387 genes in the independent cohort, which we obtained from the Gene Expression Omnibus (accession no. GSE473), and used this select set of genes as inputs to a decision tree model.[E7,E8] We used the *tree* package in R 3.0.2 to develop a binary recursive partitioning decision tree.

The tree is created by recursively partitioning the set of genes and fitting a simple prediction model within each partition. Success of the model is based on the total number of misclassifications. For example, our goal was to use a set of 501 genes, $X1, X2,\ldots,Xp$, identified within our data set to predict atopic status, Y, in an independent data set. We created our tree by repeatedly partitioning the set of genes into smaller and smaller regions of interacting genes, such that we obtained a set of predictors (genes) that we could use to identify atopic and nonatopic study subjects. Fig 3 depicts our decision tree. In this tree each node represents a cell of the partition and is associated with a model (eg, the expression value being greater than or less than a particular cutoff) that applies in that cell only. A point x belongs to a leaf if x falls in the corresponding cell of the partition. To figure out in which cell we should end up, we begin at the top (root) node of the tree and ask a series of questions about the levels of gene expression for each gene represented by the interior nodes of the tree. The questions asked will depend on the previous answers to questions at higher levels of the tree. We chose this particular method to classify study subjects as atopic or nonatopic because it provided a fast and intuitive way to classify patients that could easily be adapted to a clinical setting.

From the set of 387 inputs to the model, recursive partitioning demonstrated that 5 genes served as the optimal predictors of atopic status in the independent data set.

REFERENCES

E1. Kendziorski CM, Newton MA, Lan H, Gould MN. On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. Stat Med 2003;22:3899-914.
E2. Newton MA, Kendziorski CM, Richmond CS, Blattner FR, Tsui KW. On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. J Comput Biol 2001;8:37-52.
E3. Newton MA, Kendziorski CM. The analysis of gene expression data: methods and software. Berlin: Springer Verlag; 2003.
E4. Newton MA, Noueiry A, Sarkar S, Ahlquist P. Detecting differential gene expression with a semiparametric hierarchical mixture method. Biostatistics 2004;5:155-76.
E5. Yuan M, Kendziorski C. A unified approach for simultaneous gene clustering and differential expression identification. Biometrics 2006;62:1089-98.
E6. Shao J. Mathematical statistics. Berlin: Springer Verlag; 1999.
E7. Breiman L. Classification and regression trees. Belmont (CA): Wadsworth International Group; 1984.
E8. Ripley BD. Pattern recognition and neural networks. Cambridge (UK): Cambridge University Press; 1996.

TABLE E1. Description of gene pattern interpretations

Pattern no.	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Interpretation
1	1	1	1	1	1	Null hypothesis (none different)
2	1	2	3	4	5	All different
3	2	1	1	1	1	Cluster 1 different, others similar
4	1	2	1	1	1	Cluster 2 different, others similar
5	1	1	2	1	1	Cluster 3 different, others similar
6	1	1	1	2	1	Cluster 4 different, others similar
7	1	1	1	1	2	Cluster 5 different, others similar
8	2	2	1	1	1	Clusters 1 and 2 similar, others similar
9	2	1	2	1	1	Clusters 1 and 3 similar, others similar
10	2	1	1	2	1	Clusters 1 and 4 similar, others similar
11	2	1	1	1	2	Clusters 1 and 5 similar, others similar
12	1	2	2	1	1	Clusters 2 and 3 similar, others similar
13	1	2	1	2	1	Clusters 2 and 4 similar, others similar
14	1	2	1	1	2	Clusters 2 and 5 similar, others similar
15	1	1	2	2	1	Clusters 3 and 4 similar, others similar
16	1	1	2	1	2	Clusters 3 and 5 similar, others similar
17	1	1	1	2	2	Clusters 4 and 5 similar, others similar
18	1	2	3	1	1	Clusters 1, 4, and 5 similar, others different
19	1	1	2	3	1	Clusters 1, 2, and 5 similar, others different
20	1	1	1	2	3	Clusters 1, 2, and 3 similar, others different
21	2	3	1	1	1	Clusters 3, 4, and 5 similar, others different
22	2	2	3	1	1	Clusters 1 and 2, similar, clusters 4 and 5 similar
23	2	2	1	3	1	Clusters 1 and 2, similar, clusters 3 and 5 similar
24	2	2	1	1	3	Clusters 1 and 2, similar, clusters 3 and 4 similar
25	2	3	2	1	1	Clusters 1 and 3, similar, clusters 4 and 5 similar
26	2	1	2	3	1	Clusters 1 and 3, similar, clusters 2 and 5 similar
27	2	1	2	1	3	Clusters 1 and 3, similar, clusters 2 and 4 similar
28	2	1	1	2	3	Clusters 1 and 4, similar, clusters 2 and 3 similar
29	2	3	1	2	1	Clusters 2 and 4, similar, clusters 3 and 5 similar
30	2	1	3	2	1	Clusters 2 and 4, similar, clusters 2 and 5 similar
31	2	1	1	3	2	Clusters 1 and 5, similar, clusters 2 and 3 similar
32	2	1	3	1	2	Clusters 1 and 5, similar, clusters 2 and 4 similar
33	2	3	1	1	2	Clusters 1 and 5, similar, clusters 3 and 4 similar
34	1	1	2	3	4	Clusters 1 and 2, similar, others different
35	1	2	1	3	4	Clusters 1 and 3, similar, others different
36	1	2	3	1	4	Clusters 1 and 4, similar, others different
37	1	2	3	4	1	Clusters 1 and 5, similar, others different
38	2	1	1	3	4	Clusters 2 and 3, similar, others different
39	2	1	3	1	4	Clusters 2 and 4, similar, others different
40	2	1	3	4	1	Clusters 2 and 5, similar, others different
41	2	3	1	1	4	Clusters 3 and 4, similar, others different
42	2	3	1	4	1	Clusters 3 and 5, similar, others different
43	2	3	4	1	1	Clusters 4 and 5, similar, others different
44	1	1	2	1	3	Clusters 1, 2, and 4 similar, others different
45	1	2	1	3	1	Clusters 1, 3, and 5 similar, others different
46	1	2	1	1	3	Clusters 1, 3, and 4 similar, others different
47	2	1	3	1	1	Clusters 2, 4, and 5 similar, others different
48	2	1	1	3	1	Clusters 2, 3, and 5 similar, others different
49	2	1	1	1	3	Clusters 2, 3, and 4 similar, others different

TABLE E2. Distribution of traits across original phenotypic clusters

	Cluster 1 (n = 300)	Cluster 2 (n = 202)	Cluster 3 (n = 218)	Cluster 4 (n = 225)	Cluster 5 (n = 96)	*P*
AOE classification:	**LLL**	**HLL**	**HHM**	**MHH**	**HHH**	**value**
Asthma history						
Age (y)	3.52 ± 2.63	3.09 ± 2.40	3.66 ± 2.62	2.21 ± 1.89	2.27 ± 1.80	<.001
Total hospitalized for asthma (%)	0 (0)	0 (0)	1 (0.46)	225 (100)	94 (97.9)	<.001
ED visits for asthma (visits/100 person-years)	44.3	47.0	70.2	75.6	101	<.001
Atopic features						
History of atopic dermatitis (%)	0 (0)	202 (100)	2 (0.1)	0 (0)	94 (97.9)	<.001
History of hay fever (%)	61 (20.3)	132 (65.3)	191 (87.6)	119 (52.9)	54 (56.3)	<.001
History of positive skin test result (%)	230 (76.7)	185 (91.6)	209 (95.9)	198 (88)	92 (95.8)	<.001
Total serum IgE levels ($\log_{10}$)	2.37 ± 0.70	2.72 ± 0.72	2.79 ± 0.58	2.64 ± 0.61	2.81 ± 0.63	<.001
Spirometry						
Prebronchodilator FEV_1 (% predicted)	96.4 ± 12.7	97.7 ± 14.8	89.7 ± 13.9	91.4 ± 13.8	92.0 ± 16.1	<.001
Prebronchodilator FEV_1/FVC ratio (% predicted)	81.8 ± 7.68	81.5 ± 7.59	77.6 ± 8.54	77.8 ± 8.24	78.6 ± 9.60	<.001
Prebronchodilator peak flow	276.1 ± 67.3	274.3 ± 73.3	276.7 ± 69.1	276.4 ± 70.8	255.6 ± 73.3	.12
Airway responsiveness						
Methacholine PC_{20} (natural log)	0.71 ± 1.03	0.14 ± 1.11	−0.54 ± 1.00	0.038 ± 1.14	−0.23 ± 1.17	<.001
FEV_1 bronchodilator response (L)	0.077 ± 0.07	0.097 ± 0.08	0.12 ± 0.11	0.12 ± 0.11	0.16 ± 0.14	<.001
Anthropomorphic features						
BMI (kg/m^2)	18.1 ± 3.46	18.6 ± 3.83	18.5 ± 3.66	17.8 ± 3.19	17.6 ± 3.38	.07
Waist/hip ratio	0.882 ± 0.06	0.885 ± 0.07	0.881 ± 0.06	0.874 ± 0.05	0.877 ± 0.07	.80
Peripheral blood counts						
Eosinophils ($\log_{10}$)	2.35 ± 0.55	2.54 ± 0.53	2.57 ± 0.52	2.50 ± 0.49	2.71 ± 0.41	<.001
Lymphocytes (%)	42.1 ± 11.5	40.9 ± 9.82	41.1 ± 10.9	41.7 ± 10.5	40.8 ± 9.78	.67
Neutrophils (%)	45.5 ± 12.2	44.5 ± 11.0	44.6 ± 11.5	44.7 ± 11.5	43.1 ± 11.0	.62

We constructed an Atopy-Obstruction-Exacerbation (AOE) scoring scheme, assigning low (L), medium (M), or high (H) scores for each factor group. *ED*, Emergency department; *FVC*, forced vital capacity.

TABLE E3. Genes enriched for GR TFBSs

Symbol	Description
ADD3	Adducin 3 (gamma)
AGL	Amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase
ALPP	Alkaline phosphatase, placental
ATM	Ataxia telangiectasia mutated
ATP1B3	ATPase, Na+/K+ transporting, beta 3 polypeptide
ATP6V0C	ATPase, H+ transporting, lysosomal 16kDa, V0 subunit c
CARS	Cysteinyl-tRNA synthetase
CDC42	Cell division cycle 42 (GTP binding protein, 25kDa)
AP2M1	Adaptor-related protein complex 2, mu 1 subunit
COMT	Catechol-O-methyltransferase
DR1	Downregulator of transcription 1, TBP-binding (negative cofactor 2)
GSK3B	Glycogen synthase kinase 3 beta
GTF2H1	General transcription factor IIH, polypeptide 1, 62kDa
GUSB	Glucuronidase, beta
HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2/B1
HNRNPL	Heterogeneous nuclear ribonucleoprotein L
ITGB1	Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)
JAK1	Janus kinase 1
KPNA3	Karyopherin alpha 3 (importin alpha 4)
M6PR	Mannose-6-phosphate receptor (cation dependent)
MSH2	Mdm4 p53 binding protein homolog (mouse)
MDM4	mutS homolog 2, colon cancer, nonpolyposis type 1 (E coli)
PPP1R12A	Protein phosphatase 1, regulatory subunit 12A
NUBP1	Nucleotide binding protein 1
RPL10A	Ribosomal protein L10a
NFX1	Nuclear transcription factor, X-box binding 1
P4HA1	Prolyl 4-hydroxylase, alpha polypeptide I
PCNA	Proliferating cell nuclear antigen
PDE8A	Phosphodiesterase 8A
PDK1	Pyruvate dehydrogenase kinase, isozyme 1
PMS1	PMS1 postmeiotic segregation increased 1 *(S cerevisiae)*
CTSA	Cathepsin A
PRKCI	Protein kinase C, iota
MAP2K2	Mitogen-activated protein kinase 2
PSMA7	proteasome (prosome, macropain) subunit, alpha type, 7
RANBP2	RAN binding protein 2
RAP1B	RAP1B, member of RAS oncogene family
RPS6KB2	Ribosomal protein S6 kinase, 70kDa, polypeptide 2
MSMO1	Methylsterol monooxygenase 1
SMARCC2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 2
UAP1	UDP-N-acteylglucosamine pyrophosphorylase 1
SRM	Spermidine synthase
TCEA1	Transcription elongation factor A (SII), 1
TCF12	Transcription factor 12
TRAPPC10	Trafficking protein particle complex 10
UGP2	UDP-glucose pyrophosphorylase 2
YWHAG	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide
ZBTB17	Zinc finger and BTB domain containing 17
MAPKAPK3	Mitogen-activated protein kinase-activated protein kinase 3
BAP1	BRCA1 associated protein-1 (ubiquitin carboxy-terminal hydrolase)
PEX3	Peroxisomal biogenesis factor 3
PRKRA	Protein kinase, interferon-inducible double stranded RNA dependent activator
EIF3A	Eukaryotic translation initiation factor 3, subunit A
RPL14	Ribosomal protein L14
SLC33A1	Solute carrier family 33 (acetyl-CoA transporter), member 1
VAPA	VAMP (vesicle-associated membrane protein)–associated protein A, 33kDa
CNOT8	CCR4-NOT transcription complex, subunit 8
SEP15	Selenoprotein isoform 1 precursor
ZRANB2	Zinc finger, RAN-binding domain containing 2
MED23	Mediator complex subunit 23
SPAG7	Sperm associated antigen 7
PREPL	Prolyl endopeptidase-like

(Continued)

TABLE E3. *(Continued)*

Symbol	Description
VPS9D1	Chromosome 16 open reading frame 7
HS2ST1	Heparan sulfate 2-O-sulfotransferase 1
FLOT1	Flotillin 1
BCKDK	Branched chain ketoacid dehydrogenase kinase
CITED2	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2
NDC80	NDC80 kinetochore complex component homolog *(S cerevisiae)*
TAB1	TGF-beta activated kinase 1/MAP3K7 binding protein 1
MTF2	Metal response element binding transcription factor 2
SACM1L	SAC1 suppressor of actin mutations 1-like (yeast)
ADNP	Activity-dependent neuroprotector homeobox
SIRT1	sirtuin 1
PHF3	PHD finger protein 3
NIPBL	Nipped-B homolog (Drosophila)
WSB1	WD repeat and SOCS box containing 1
NARF	nuclear prelamin A recognition factor
RANBP6	RAN binding protein 6
AKAP8L	A kinase (PRKA) anchor protein 8-like
NPTN	neuroplastin
ALG5	Asparagine-linked glycosylation 5, dolichyl-phosphate beta-glucosyltransferase homolog *(S cerevisiae)*
PNPLA8	Patatin-like phospholipase domain containing 8
VPS36	Vacuolar protein sorting 36 homolog *(S cerevisiae)*
GLOD4	Glyoxalase domain containing 4
HSD17B11	Hydroxysteroid (17-beta) dehydrogenase 11
NT5C3	5′-Nucleotidase, cytosolic III
MRPL30	Mitochondrial ribosomal protein L30
DDX41	DEAD (Asp-Glu-Ala-Asp) box polypeptide 41
NLK	nemo-like kinase
LUC7L3	LUC7-like 3 (S. cerevisiae)
BTBD1	BTB (POZ) domain containing 1
RAB4B	RAB4B, member RAS oncogene family
SLC38A2	Solute carrier family 38, member 2 family with sequence similarity 35, member A
FAM35A	Late endosomal/lysosomal adaptor, MAPK and MTOR activator 1
LAMTOR1	Chromosome 19 open reading frame 24
C19orf24	Pleckstrin homology domain containing, family B (evectins) member 2
PLEKHB2	Zinc finger, CCHC domain containing 8
ZCCHC8	Kinesin family member 21A
KIF21A	Ankyrin repeat domain 10
ANKRD10	THUMP domain containing 1
THUMPD1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27
DDX27	Asunder, spermatogenesis regulator homolog (Drosphila)
ASUN	Lysine (K)-specific demethylase 3A
KDM3A	Kelch-like 7 (Drosophila)
KLHL7	Chromosome 19 open reading frame 10
C19orf10	SAR1 homolog A *(S cerevisiae)*
SAR1A	CDC42 small effector 2
CDC42SE2 UTP3	UTP3, small subunit (SSU) processome component, homolog *(S cerevisiae)*
PHTF2	Putative homeodomain transcription factor 2
SNX14	Sorting nexin 14
POLD4	Polymerase (DNA-directed), delta 4, accessory subunit
PLEKHA1	Pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1
RRAGC	Ras-related GTP binding C
IKZF5	IKAROS family zinc finger 5 (Pegasus)
ACD	Adrenocortical dysplasia homolog (mouse)
FAM65A	Family with sequence similarity 65, member A
SLTM	SAFB-like, transcription modulator
KIAA0319L	KIAA0319-like
C10orf88	Chromosome 10 open reading frame 88
IFT74	intraflagellar transport 74 homolog (Chlamydomonas)
MED25	Mediator complex subunit 25
STARD3NL	STARD3 N-terminal like
MCM8	Minichromosome maintenance complex component 8
WDR73	WD repeat domain 73

(Continued)

TABLE E3. *(Continued)*

Symbol	Description
ZNF700	Zinc finger protein 700
SLC25A51	Solute carrier family 25, member 51
G6PC3	Glucose 6 phosphatase, catalytic, 3
CASC4	Cancer susceptibility candidate 4
ESCO1	Establishment of cohesion 1 homolog 1 *(S cerevisiae)*
AGAP3	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3
ANAPC16	Anaphase promoting complex subunit 16
LRRC28	Leucine rich repeat containing 28
C17orf49	Chromatin complexes subunit BAP18
PM20D2	Peptidase M20 domain containing 2
PPP4R2	Protein phosphatase 4, regulatory subunit 2
ZNF296	Zinc finger protein 296
DENND6A	Family with sequence similarity 116, member A
C9orf72	Chromosome 9 open reading frame 72
TIPRL	TIP41, TOR signaling pathway regulator-like *(S cerevisiae)*
POC1B	POC1 centriolar protein homolog B (Chlamydomonas)
TSEN54	tRNA splicing endonuclease 54 homolog (S cerevisiae)
ZNF181	Zinc finger protein 181
C6orf120	Chromosome 6 open reading frame 120