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