

Accepted Manuscript

Strength in numbers: The quest for asthma genes

Donata Vercelli, MD, Eugene R. Bleeker, MD

PII: S0091-6749(19)30817-6

DOI: <https://doi.org/10.1016/j.jaci.2019.06.007>

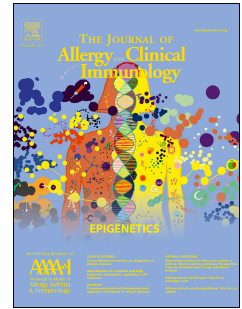
Reference: YMAI 14046

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 3 June 2019

Revised Date: 11 June 2019

Accepted Date: 17 June 2019



Please cite this article as: Vercelli D, Bleeker ER, Strength in numbers: The quest for asthma genes, *Journal of Allergy and Clinical Immunology* (2019), doi: <https://doi.org/10.1016/j.jaci.2019.06.007>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Strength in numbers: The quest for asthma genes

Donata Vercelli, MD^{1,2} and Eugene R. Bleeker, MD^{2,3}

Department of Cellular and Molecular Medicine; Asthma and Airway Disease Research Center and
Division of Genetics, Genomics and Precision Medicine, Department of Medicine, The University of
Arizona Health Sciences, Tucson, AZ, USA

Corresponding Author:
Donata Vercelli, MD
The BIO5 Institute, Rm. 339
1657 E. Helen Street
Tucson, AZ 85721
Phone: (520) 626-6387
FAX: (520) 626-6623
Email: donata@email.arizona.edu

Key words:

genetic risk factors; childhood-onset asthma; adult-onset asthma

DV is currently supported by research grants from the NIAID, the NHLBI and VIFOR/OM-Pharma and previously received research support from Johnson & Johnson. ERB has undertaken clinical trials through his employers, Wake Forest School of Medicine and University of Arizona, for AstraZeneca, MedImmune, Boehringer Ingelheim, Genentech, Novartis, Regeneron, and Sanofi Genzyme, and has served as a paid consultant for ALK-Abello, AstraZeneca, Glaxo Smith Kline, MedImmune, Novartis, Regeneron, Sanofi Genzyme, and TEVA.

Abstract

The common variants that influence the architecture of complex human diseases such as asthma have extremely small effects that can only be captured by very large studies. A group of geneticists and statisticians from the University of Chicago relied on the UK Biobank (a publicly available repository of genetic and phenotypic data collected on approximately 500,000 individuals from across the United Kingdom) to identify both shared and distinct genetic risk factors for childhood- and adult-onset asthma and thus assess whether these are the same or distinct diseases. Using a bold analytical approach that combined genome-wide association studies (GWAS), tissue enrichment patterns for expressed GWAS loci, and genotype-imputed gene expression data, the authors identified loci with both age of onset-specific effects and shared effects, and biological pathways and tissues potentially linking genetic variation to asthma pathogenesis.

9433 adults with childhood-onset asthma, 21 564 adults with adult-onset asthma, 6849 individuals with adolescent or young adult-onset asthma, and 318 237 controls, all of whom were studied not only in genome-wide but also in transcriptome-wide association studies – and all of that predicated on existing data and a computational tour de force, without ever holding a pipette or setting foot in a clinical lab (1). This is how a clever group of geneticists and statisticians from the University of Chicago addressed (and to a significant, albeit not complete, extent) solved the “need-for-big-data” problem that has been plaguing asthma and (more generally) complex disease genetics over the last decade. The common variants that appear to influence the architecture of complex human traits at the population level have extremely small effects that can only be captured by very large studies. In response to this realization, breathtakingly large data repositories built with government funds are now becoming available and changing the face of genetics as we knew it. Unlike many of the cohorts previously developed, these repositories collect unselected samples and thus are (hoped to) faithfully represent the genetic and phenotypic architecture of the original populations. One such data set is the UK Biobank, a prospective cohort study with genetic and phenotypic data collected on approximately 500,000 individuals from across the United Kingdom, aged between 40 and 69 at recruitment (2). Although it includes almost exclusively samples from individuals of white European descent and thus does not address the important question of ethnic and racial heterogeneity in disease prevalence and severity (3), the UK Biobank is currently unique among open resources for its size and scope. With palpable pride, its curators state that “a rich variety of phenotypic and health-related information is available on each participant, including biological measurements, lifestyle indicators, biomarkers in blood and urine, and imaging of the body and brain. Follow-up information is provided by linking health and medical records. Genome-wide genotype data have been collected on all participants, providing many opportunities for the discovery of new genetic associations and the genetic bases of complex traits.” By comparison, the ongoing National Heart, Lung and Blood Institute-sponsored Trans-Omics for Precision Medicine (TOPMed) program (<https://www.nhlbiwgs.org/>), which is shaped by similar goals, contains whole genome sequencing data and extensive, standardized cross-sectional and longitudinal phenotypic information, but as of October 2018 consisted “only” of ~144,000 participants with a focus on heart, lung, blood, and sleep disorders.

Thus it was that the University of Chicago group chose to rely on the UK Biobank to identify both shared and distinct genetic risk factors for childhood- and adult-onset asthma and thus assess whether these are the same or distinct diseases. There is value to the use of large data sets in common diseases such as asthma. Besides sheer numbers of samples, genotyping is usually performed in established approved central laboratories. Existing data can be analyzed by imputation, and imputed data can be further related to genomic signals. Similar to the US Million Veterans Program and All of Us, the data rely extensively on electronic health records, which allows for large sample sizes at costs that academic groups can finally afford. On the other hand, there are potential pitfalls, some of which the authors themselves acknowledge. Disease misclassification between asthma and more environmentally-induced airway diseases such as chronic obstructive pulmonary disease (COPD) occurs often. The effect of tobacco exposure blurs the distinction between asthma and COPD. Indeed, patients with Asthma-COPD Overlap (ACO) syndrome often receive an asthma diagnosis even though their disease is more compatible with COPD. The reported values for active smoking may not fully reflect actual prior tobacco exposure and thus hinder the analysis of gene-environment interactions that are important in this group of subjects. Moreover, data on gene expression – a major component of this work – were not actually measured but rather imputed using prediction models trained in large reference databases of genotypes and tissue-

specific gene expression profiles. With these genotype-imputed expression profiles in hand, gene-based association tests can be performed to correlate the predicted expression levels with phenotypes (e.g. asthma) and identify candidate causal genes from GWAS data. Although this is one of the most creative and interesting approaches used in this statistically sophisticated paper, some genes might have been misevaluated because of difficult imputation. Finally, the gene expression data used to predict candidate genes were derived from heterogeneous tissues collected mostly from adults, implying that genes with developmentally or environmentally regulated expression might have been missed.

Because all of these difficulties are real and potentially daunting, it is even more striking that of 61 loci found to be associated with asthma (56 in childhood-onset and 19 in adult-onset disease), as many as 33 actually replicated previous findings, including the chromosome 17q locus as the most significant signal for childhood-onset asthma and HLA as the region most significantly associated with adult-onset asthma. Overall, these results powerfully validate the analytical strategies chosen in this study and their ability to tease out of a seemingly unwieldy data set genetic information that makes sense. In turn, this validation implies that the asthma-associated loci reported by this study for the first time (**Figure 1**) should be further explored - both the more "obvious" ones, such as *IRF4*, *CCL20*, *IL6*, *MUC5AC*, *FADS2*, *TBX21* and *RUNX1*, and those more unexpected, such as cytokine and chemokine receptors and signaling molecules. It is also worth noting that essentially all the new asthma genes identified in this study have relatively modest genome-wide significance and therefore fall within the "medium/high hanging fruit" group (4) that may have been made discoverable primarily by the high statistical power of this unusually large analysis.

And there is more. By linking expression of genes mapping to childhood-onset or adult-onset asthma loci to specific tissues, the authors demonstrated that genes at childhood-onset loci were most highly expressed in skin, whereas genes at adult-onset loci were most highly expressed in lung and spleen. These findings provide independent support for the notion that the inception of childhood asthma is triggered by defects in barrier function that alter mucosal immune responses and disturb the mucosal/microbial interface. These defects appear to have strong genetic determinants. On the other hand, the enrichment for lung-and spleen-expressed genes in adult-onset asthma (a condition that shares many genetic risk factors with childhood-onset asthma but with overall much smaller effects) points to organ-centered immune events, possibly triggered by environmental exposures.

Studies of this kind bridge past, present and future. They confirm what we knew and offer a platform to identify novel candidates and test intriguing hypotheses. For instance, studies with sputum, airway brushings and broncho-alveolar lavage, especially in adults, will allow to both replicate and extend these results before testing them in other biological models (5, 6). Large data sets might provide the granularity required to begin formulating testable hypotheses about the pathways underlying adult-onset asthma, a disease that has a reduced or absent family history and less active type-2 immune mechanisms, yet may be neutrophilic or eosinophilic (7). It will also be critical to understand the genetic influences involved in (biological) sex effects and patterns of asthma severity in childhood and later in life. Recent studies show that more boys than girls, and more women than men, have asthma; genetic as well as hormonal factors may underlie these differences. Disease severity and progression also differ sharply among ethnic groups; large studies in multi-ethnic populations will be essential to illuminate these clinically relevant issues (7, 8), support precision pharmacogenetics approaches (9) and deconvolute gene-environment

interactions potentially underpinning racial disparities. But the road ahead will be long and possibly bumpy. More comprehensive standardized sub-phenotyping will be necessary to dissect the different gene and environmental factors that cause heterogeneity in this complex disease. Indeed, even in a study as large as the University of Chicago one, genome-wide significant loci accounted for only 32.7% of the heritability of childhood-onset asthma, and only 9.8% of the heritability of adult-onset asthma. This suggests that many more asthma loci remain to be discovered, but their very small effects will be detectable only in even larger cohorts – which means that this may be the end of the beginning, rather than the beginning of the end, of asthma genetics.

1. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med*. 2019;7(6):509-22.
2. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-9.
3. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat Commun*. 2016;7:12522.
4. Ober C. Asthma genetics in the post-GWAS era. *Ann Am Thorac Soc*. 2016;13 Suppl 1:S85-90.
5. Li X, Hastie AT, Hawkins GA, Moore WC, Ampleford EJ, Milosevic J, et al. eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. *Allergy*. 2015;70(10):1309-18.
6. Modena BD, Bleecker ER, Busse WW, Erzurum SC, Gaston BM, Jarjour NN, et al. Gene Expression Correlated with Severe Asthma Characteristics Reveals Heterogeneous Mechanisms of Severe Disease. *Am J Respir Crit Care Med*. 2017;195(11):1449-63.
7. 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(4):315-23.
8. 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*. 2010;127(2):382-9 e13.
9. Ortega VE, Meyers DA, Bleecker ER. Asthma pharmacogenetics and the development of genetic profiles for personalized medicine. *Pharmgenomics Pers Med*. 2015;8:9-22.

Legends

Figure 1. **Miami plot of childhood-onset (top graph, in blue) and adult-onset (bottom graph, in red) asthma GWAS results from Ref.(1).** Each point corresponds to a single nucleotide polymorphism. The x axis shows the position of each polymorphism along the 22 autosomes; the y axis shows the $-\log_{10}$ p values from the childhood-onset GWAS and the adult-onset GWAS. The red dashed line indicates the genome-wide significance threshold (5×10^{-8}). The boxes highlight childhood-onset asthma-specific (blue), adult-onset asthma-specific (dashed red), and shared (blue/red) loci not previously reported in GWAS (adapted from Ref. 1).

