

Accepted Manuscript

Identification of novel allergic diathesis genes: are we closer to novel therapeutic targets?

Michelle Daya, PhD, Kathleen C. Barnes, PhD



PII: S0091-6749(18)31739-1

DOI: <https://doi.org/10.1016/j.jaci.2018.11.028>

Reference: YMAI 13769

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

Received Date: 25 May 2018

Revised Date: 15 November 2018

Accepted Date: 23 November 2018

Please cite this article as: Daya M, Barnes KC, Identification of novel allergic diathesis genes: are we closer to novel therapeutic targets?, *Journal of Allergy and Clinical Immunology* (2019), doi: <https://doi.org/10.1016/j.jaci.2018.11.028>.

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.

1 Identification of novel allergic diathesis genes: are we closer to novel therapeutic 2 targets?

3
4 **Michelle Daya, PhD and Kathleen C. Barnes, PhD**

5
6 Department of Medicine, University of Colorado Denver, Aurora, CO 80045

7 8 **Corresponding author details**

9 Dr. Kathleen C. Barnes

10 University of Colorado

11 Colorado Center for Personalized Medicine

12 Bldg 500, Mail Stop F563

13 Anschutz Medical Campus

14 13001 E 17th Place, Room 5330A, 5th Fl East

15 Phone: 303-724-9734

16 Email: KATHLEEN.BARNES@UCDENVER.EDU

17
18 Key words: genome wide association study (GWAS), allergic diathesis, expression
19 quantitative trait loci (eQTL), team science

20
21 Disclosure of potential conflict of interest: The authors declare that they have no
22 relevant conflicts of interest.

23
24
25 Until recently, asthma and allergic disease genetics research followed a conventional
26 approach dating back to the early 2000's, whereby a genome-wide association study
27 (GWAS) was performed on a single discovery sample followed by replication in an
28 independent sample, and ideally (but not always) followed up with a validation study of
29 the top gene or genes to support biological relevance of the candidate(s). Through
30 GWAS, >28,000 variants associated with >1,800 different phenotypes and diseases
31 have been discovered to date(1). For asthma alone, more than two dozen GWAS have
32 been performed(2) yielding associations that have proven to be highly replicable and
33 potentially 'true positives'(2, 3).

34 The bad news is that, for complex traits (i.e., allergic disease), even the most
35 significantly associated loci typically have small effect sizes and only explain a small
36 fraction of the predicted genetic variance. Moreover, with association signals spread
37 across the genome and typically near multiple genes devoid of a clear connection to
38 disease, it is often difficult to pinpoint the biological processes underlying the observed
39 associations. Allergic disease poses the added challenge of being a complex trait for
40 which diagnostic criteria do not rely on biological markers. Disentangling the genetic
41 basis for eczema, hay fever, and asthma has also proven difficult given that these traits
42 often present sequentially in the life of the patient (the 'atopic march'), are often
43 comorbid, and the development of one phenotype can predict the development of
44 another(4). Regardless of the debate whether manifestation of eczema, hay fever and
45 asthma is a result of a causal relationship or a shared cause, genetic findings support
46 both common and unique pathways, and the study by Ferreira and colleagues in this
47 issue of the *Journal of Allergy and Clinical Immunology* not only identifies additional risk
48 loci for allergic disease, but identifies genes whose inhibition may lead to the
49 identification of novel therapeutic targets for the treatment of allergic disease(5).

50 A major challenge with traditional GWAS is the identification of causal genetic
51 variants that in turn may lead to biological insight. A novel feature of the report by

Ferreira and colleagues is that this limitation is somewhat overcome by employing a new gene-based approach EUGENE. EUGENE identifies genes associated with disease by combining evidence of expression quantitative trait loci (eQTLs) associated with gene expression in one or more tissues, easing biological interpretation. By using composite evidence from multiple independent genetic variants regulating gene expression, statistical power to detect genes involved in disease pathogenesis is increased. This is not a new idea *per se*, however, as association signals are typically aggregated based on functionally significant coding variants within a gene. Because genetic variants identified by GWAS are often enriched in regulatory regions of the genome, as recently highlighted in the largest asthma meta-analysis to date(3), aggregation based on coding variants may have limited utility, and the EUGENE approach to aggregation may be more relevant in understanding complex disease. While all of these methods have their strengths and weaknesses and operate under different assumptions, EUGENE provides a complementary and powerful approach to discover additional genetic underpinnings of disease.

Demonstrating the utility of this approach, the report by Ferreira et al identifies a number of novel genes common to asthma, hay fever and eczema. These genes were not detected in their previous GWAS of the common pathogenesis for the same traits (Figure 1). In addition, a number of the discoveries involve plausible candidate genes with functions directly relevant to allergic disease. The authors further demonstrate the robustness of their results by successfully replicating 20 of the 30 identified gene associations in independent data from the robust UK Biobank dataset of over 233,000 individuals. Four of the replicated genes involved genetically controlled *decreased* expression in healthy subjects, and are compelling targets for the development of therapeutic interventions that inhibit the proteins coded by the genes. In addition, the identification of allergy protective eQTLs associated with changes in gene expression (e.g. allergy protective eQTLs for the *FOSL2* gene increase expression of the gene), may also yield new drug discoveries. Although functional work is needed to confirm a causal link between these findings and allergy risk, the authors rightfully contend that these discoveries may lead to new therapeutic interventions.

The successful identification of novel genetic associations for allergic disease in this study is emblematic of a new generation of GWAS and enabled by two major (and positive) shifts in GWAS research. First, the Ferreira et al report owes its success, in part, to the notion of 'recycling' existing GWAS data, not only by applying novel analytical strategies (e.g., EUGENE) but also by leveraging multiple existing datasets (e.g., 13 studies comprising >360,000 patients) through collaborative team science. Indeed, it has been argued for several years that fundamental challenges in GWAS including inadequate sample sizes and phenotype heterogeneity can be overcome by establishing collaborations and consortia in which investigators share GWAS data to achieve adequate power and increase power to study less common outcomes or secondary phenotypes. In the midst of the vigorous debate over data deposition and the broad sharing of data for the public good(6), consistent with funding policies(7), there is a shift from emphasizing attributions by individual researchers to sharing resources across large consortia and to the broader community, as has been the case for asthma for nearly a decade, with the establishment of large national (EVE(8), GABRIEL(9)) and multi-national consortia (CAAPA(10), TAGC(3)). Another shift is that summary association statistics are increasingly being deposited online as part of GWAS publications (e.g., hundreds of such datasets are available), and are also increasingly available through efforts such as the UK BioBank. In addition to leveraging summary association statistics, EUGENE utilizes summary eQTL statistics that are also becoming increasingly available through efforts such as GTEx. With this advancement, insights

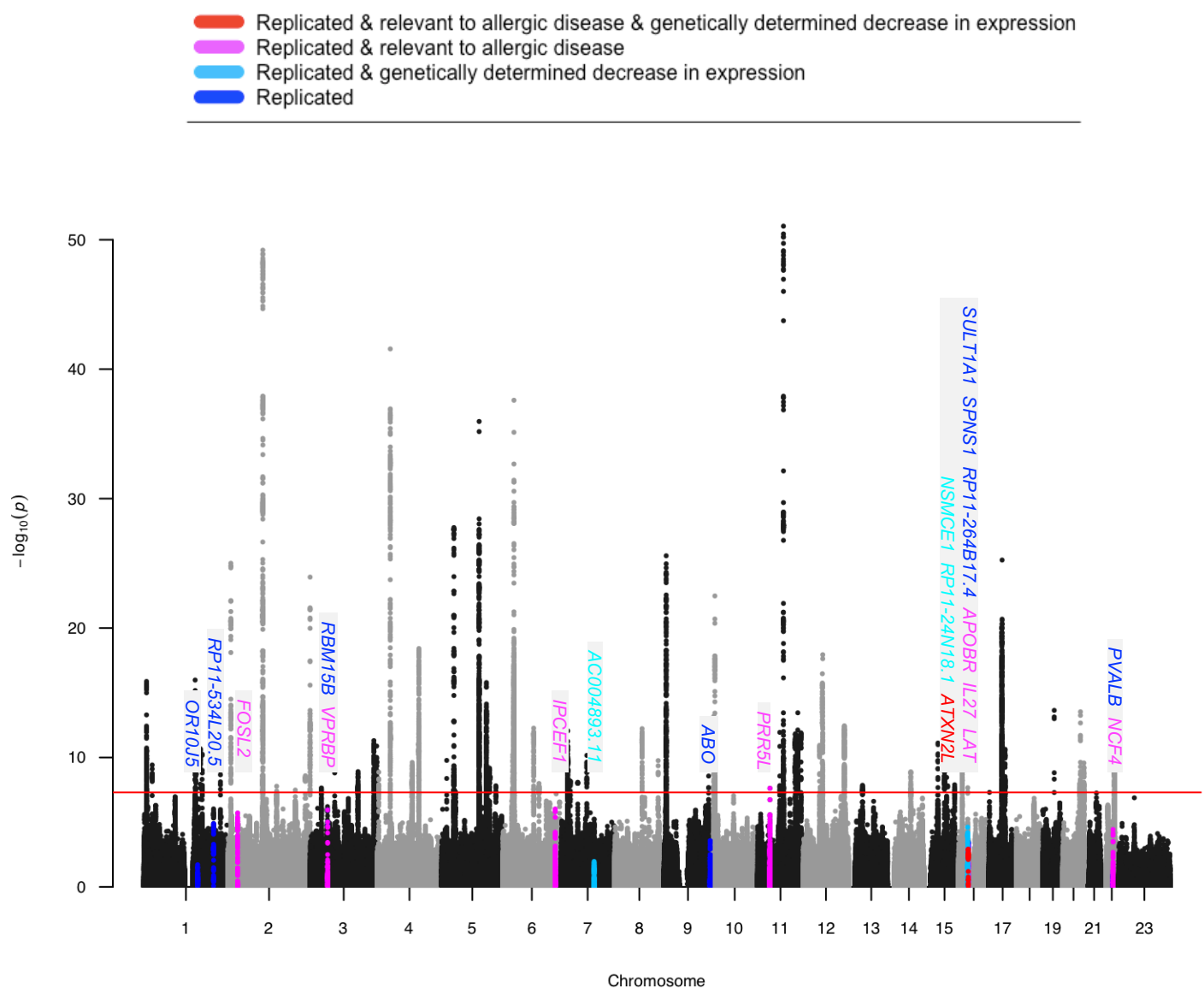
103 into disease pathogenesis can be gained without requiring additional costly generation of
104 transcriptomic data.

105 A shortcoming of this report includes a focus on European ancestry cohorts,
106 limiting the generalizability of these findings to other ethnic and racial groups. Sadly, this
107 handicap is owed in large part to the dearth of representative GWAS's focused on non-
108 European populations, and elsewhere this issue has proven equally challenging(3). The
109 authors also acknowledge that robust functional studies are essential for drawing firm
110 conclusions regarding the contribution of the 19 identified genes toward allergic disease
111 pathology. The EUGENE approach only leverages information on eQTLs, and as many
112 disease associations are co-localized with DNA methylation or histone modification
113 QTLs that may not be eQTLs, incorporating information on epigenetic QTLs may
114 contribute to additional discoveries in the future. This study only focused on the genetic
115 underpinnings of allergic disease, and the integration of environmental exposures highly
116 relevant to risk for allergy with genetic information would also be of value. Despite
117 these issues, the work of Ferreira and colleagues demonstrates how their novel gene-
118 base approach can empower robust new discoveries that provide insight into biology,
119 and can potentially lead to the development of therapeutic interventions.
120

REFERENCES

1. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 2017;45(D1):D896-D901.
2. Vicente CT, Revez JA, Ferreira MAR. Lessons from ten years of genome-wide association studies of asthma. *Clin Transl Immunology.* 2017;6(12):e165.
3. Demenais F, Margartite-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet.* 2018;50(1):42-53.
4. Barnes KC. Genes and Atopic Phenotypes. In: Wahn U, Sampson HA, editors. *Allergy, Immunity and Tolerance in Early Childhood: The First Steps of the Atopic March*; Elsevier Inc; 2016. p. 119-31.
5. Ferreira MAR, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Eleven loci with new reproducible genetic associations with allergic disease risk. *J Allergy Clin Immunol.* 2018.
6. Warren E. Strengthening Research through Data Sharing. *N Engl J Med.* 2016;375(5):401-3.
7. Paltoo DN, Rodriguez LL, Feolo M, Gillanders E, Ramos EM, Rutter JL, et al. Data use under the NIH GWAS data sharing policy and future directions. *Nat Genet.* 2014;46(9):934-8.
8. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet.* 2011;43(9):887-92.
9. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med.* 2010;363(13):1211-21.
10. 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.

Figure 1: GWAS manhattan plot of the shared genetic origin of asthma, hay fever and eczema in subjects. The plot summarizes single SNP association statistics from the previous meta-analysis and illustrate the strong association peaks that can be obtained through team science and large sample sizes. Before applying the EUGENE method, a conditional analysis was performed to remove the main effects of the previous meta-analysis, in order to discover additional genetic underpinnings of allergy. The location of the new genes discovered by EUGENE is highlighted. Only one of the gene regions has a single SNP association that crosses genome-wide significance.



The plot was generated using single SNP association summary statistics downloaded from ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/FerreiraMA_29083406_GCST005038. Gene regions of the 20 genes discovered by EUGENE are colored (+/- 10 KB). The red horizontal line denotes the genome-wide significance threshold. In red is the *ATXN2L* gene region, which replicated, is relevant to allergic disease, and whose genetically determined gene expression is protective against allergic disease. In purple are the gene regions of the remaining 8 genes that replicated and that are relevant to allergic disease. In light blue are the regions of 3 genes that replicated and whose genetically determined gene expression is protective against allergic disease, but whose function in allergic disease is not known. In blue are the gene regions of the remaining 8 genes that replicated.