

Accepted Manuscript

Association of ST2 polymorphisms with atopy, asthma and leukemia

Melissa H. Bloodworth, Ph.D., Mark Rusznak, Lisa Bastarache, M.S., Janey Wang, M.S., Joshua C. Denny, M.D., M.S., R. Stokes Peebles, Jr., M.D.



PII: S0091-6749(18)30712-7

DOI: [10.1016/j.jaci.2018.03.020](https://doi.org/10.1016/j.jaci.2018.03.020)

Reference: YMAI 13433

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

Received Date: 2 August 2017

Revised Date: 1 February 2018

Accepted Date: 19 March 2018

Please cite this article as: Bloodworth MH, Rusznak M, Bastarache L, Wang J, Denny JC, Peebles Jr. RS, Association of ST2 polymorphisms with atopy, asthma and leukemia, *Journal of Allergy and Clinical Immunology* (2018), doi: 10.1016/j.jaci.2018.03.020.

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 Association of ST2 polymorphisms with atopy, asthma and leukemia

2
3
4 Running Title: ST2 associates with allergic disease and leukemia

5
6
7 Melissa H. Bloodworth, Ph.D. ^{a*}

8 Mark Rusznak ^{b*}

9 Lisa Bastarache, M.S. ^c

10 Janey Wang, M. S. ^c

11 Joshua C. Denny, M.D., M.S. ^c

12 R. Stokes Peebles, Jr., M.D. ^{a,b}

13
14 ^aDepartment of Pathology, Microbiology, and Immunology, Vanderbilt University School of
15 Medicine, Nashville, TN, USA

16 ^bDivision of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt
17 University School of Medicine, Nashville, TN, USA

18 ^cDepartment of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN,
19 USA

20 *These authors contributed equally to this work

21
22
23 Address correspondence to: Stokes Peebles, M.D.
24 Division of Allergy, Pulmonary and Critical Care Medicine
25 Vanderbilt University School of Medicine
26 T-1218 MCN, VUMC
27 1161 21st Ave South
28 Nashville, TN 37232-2650, USA
29 Telephone: 615-343-3412; fax: 615-343-7448
30 E-mail: stokes.peebles@vanderbilt.edu

31
32 Funding:

33 R01 AI 124456 – R. S. P.

34 U19 AI 095227 – R. S. P.

35 R01 AI 111820 – R. S. P.

36 2I01BX000624 – R. S. P.

37 T32 GM07347 – Vanderbilt MSTP

38 F30 AI118376 – M. H. B.

39 ULTR000445 – Vanderbilt CTSA

40 R01 LM 010685 – J. C. D.

41
42
43 Word Count: 1,378

To the editors:

The phenome-wide association study (PheWAS) is a new, validated, reverse genetics approach that associates genetic variants of interest with phenotypes by linking a database of de-identified genotyping to a broad range of electronic medical record (EMR)-derived clinical phenotypes.¹ While the more common genome-wide association study (GWAS) is useful in identifying genetic polymorphisms that may relate to one particular disease phenotype, the PheWAS's ability to evaluate these polymorphisms against an entire database of clinical phenotypes greatly expands the scope of genomic association research. Using a PheWAS, one is able to examine polymorphisms detected in the context of a disease to not only confirm this established association with greater certainty, but one may uncover links to entirely new phenotypes that had not been previously considered or investigated at the genetic level.¹

IL-33 is one of the most consistently associated gene candidates for asthma identified by GWASs in diverse ethnic groups.² IL-33 is a central mediator of both innate and adaptive immunity regulated allergic inflammation in the lung.^{3,4} The primary binding partner of IL-33, ST2, is a member of the IL-1 receptor family that is expressed on a number of immune cells, including ILC2, CD4+ Th2 cells, mast cells, and eosinophils.⁵ GWASs investigating polymorphisms related to asthma have identified a number of single nucleotide polymorphisms (SNPs) in the ST2 receptor gene.^{6,7} We sought to investigate these polymorphisms to assess the role of ST2 signaling in other, non-asthma related allergic diseases. In addition, we hoped to identify new diseases in which genetic links to the ST2 receptor had not been reported. We performed a PheWAS on 15 ST2 genetic variants implicated in asthma to assess their associations with other allergic, and non-allergic diseases. The Exome BeadChip that we used to perform the PheWAS genotyped 38 variants in *IL1R1*. We only included variants with a minor allele threshold (MAF) of greater than 1%; this filtering is necessary to support a PheWAS analysis using a logistic regression. The PheWAS utilized a cohort of 36,400 adult (age > 18

years) subjects who had genotyping on Illumina HumanExome BeadChip version 1.1 and available electronic health record (HER) data from the Vanderbilt BioVU DNA biobank that was ascertained previously for five different criteria: (1) eligibility as case or control in one of 31 pharmacogenetics studies (2) available longitudinal data with primary care visits (3) presence in the cancer registry (4) old age with longitudinal data (5) presence of rare diseases or conditions ascertained via billing codes. Of the 15 SNPs, six had mean allele frequencies (MAFs) greater than 1% in the total cohort population, rendering them sufficiently powered for further analysis.¹ All of these six SNPs are in the ST2 gene (human chromosome 2q11.2).⁶ The PheWAS yields results as a list of phenotype ICD9 codes that best associated with each particular SNP.

Our study revealed that the minor allele of ST2 SNP rs1041973 associated with allergic conjunctivitis (OR=1.27), an allergic disease that had not been linked genetically in the context of IL-33/ST2 signaling. We further demonstrate that the minor allele of rs1041973 exhibits an inverse association with eosinophilia, as indicated by an odds ratio (ORs) < 1 (0.686, Table 1). In addition, the PheWAS confirmed the previously reported inverse relationship of 5 ST2 SNPs with asthma (rs10192036 OR=0.876, rs1041973 OR=0.505, rs4988956 OR=0.876, rs10192157 OR=0.876, rs10206753 OR=0.876), and it confirmed the formerly identified association of rs1420101 (OR=1.20) with asthma.^{6,7} We calculated the pair-wise LD for each of the 6 variants tested and found that the minor alleles of rs10192036, rs10206753, rs4988956, and rs10192157 are in perfect linkage ($r^2=1$). For this reason, the associations of these 4 SNPs are reported under only rs4988956. These genomic investigations with study populations entirely separate from our own provide ORs in similar directions for all six SNPs investigated in this study, confirming the previously described association of each.^{6,7} This further supports the validity of our method by demonstrating that genetic and phenotypic data from Vanderbilt University Medical Center patients is representative of national and international populations.⁶ Finally, the PheWAS uncovered a novel association of 2 SNPs in the ST2 gene with various forms of

leukemia (Table 2). The minor allele of the SNP rs1420101 associated with leukemia and acute myeloid leukemia, while the minor allele of rs1041973 associated with leukemia, acute myeloid leukemia, and acute lymphoid leukemia.

The results of this study inform our understanding of the ST2 pathway as it relates to other, non-asthmatic forms of allergic disease. We are also the first to identify rs1041973 as a genetic variant that associates inversely with eosinophilia. In an analysis of peripheral blood eosinophilia for eosinophilia-associated rs1041973 cases and controls, there was a significant increase in the maximum recorded percentage of peripheral blood eosinophils in cases (13.5%) compared to controls (3.4%, $p = 3.15 \times 10^{-165}$, Supplemental Table 1). Since SNPs examined in this study were originally identified in a GWAS as relevant to asthma, the associations that we report here with asthma serve as another indicator of the reliability of PheWAS as an analytical technique, and of the validity of the asthma-ST2 connection. We repeated our study in an independent cohort of over 500,000 subjects, the UK Biobank.⁸ We verified our previously identified allergic and asthmatic phenotype associations with ST2 polymorphisms in the UK Biobank using self-reported phenotype codes (Supplemental Table 2). Finally, our findings establish the first genetic link between the ST2 gene and leukemia. We were unable to confirm our associations to lymphoma and leukemia due to an insufficient number of these disease cases in the UK Biobank cohort. However, the association between ST2 polymorphisms and multiple forms of leukemia can be supported by a handful of *in vivo* and *in vitro* studies that have reported IL-33 and ST2 to be potential mediators of leukemic cell proliferation.⁹ These findings identify novel genetic risk factors for leukemia, and provide further support for the investigation of this signaling pathway in leukocyte tumorigenesis. The significance of these results is further demonstrated by the fact that 5 of the 6 SNPs discussed in this study have MAFs over 37%. These polymorphisms are present in a substantial portion of the United States population, and their continued investigation will no doubt yield widespread benefit.

While PheWASs such as this one offer inexpensive and efficient insight into the association of disease with specific genetic variants, the model itself has certain intrinsic limitations. Firstly, a selection bias exists, as the genotyping information available for the study comes from patients who have had blood drawn for laboratory testing. Additionally, a majority of the subjects who provided blood were admitted to the emergency department or hospital, further compromising the true randomness of the test samples. Furthermore, the study reports uncorrected P values as well as P values adjusted with the false discovery rate correction. A Bonferroni correction may be used to reduce the likelihood of Type I statistical error. However, this study was performed as an exploratory analysis. Therefore, we excluded the use of multiple testing corrections to reduce our Type II error rate. In situations such as this study where the number of tests is large, multiple testing corrections can be overly conservative, possibly contributing to missed or overlooked results. A limitation of our approach is that it depends on the prior identification of SNPs in other studies or the sufficiently high prevalence of SNPs in our cohort of study. For this reason, the number of SNPs available for PheWAS is limited, and important genetic associations in the ST2 gene likely remain undiscovered. In addition, inherent limitations exist in the study's dependence on disease phenotype diagnosis with ICD codes. ICD-9 billing codes are often a poor and unspecific diagnostic tool, and variation exists in the application of these codes by physicians and the ability of the algorithms to assign code designations to more narrative, textual diagnoses. This is particularly true in the case of asthma in which the diagnoses are difficult to confirm in the EHR, as not every patient will have lung function or methacholine challenge tests performed.

In conclusion, these data demonstrate the utility of PheWAS by identifying links between polymorphisms in the ST2 gene and multiple forms of leukemia. Further, they implicate specific

ST2 SNPs in allergic diseases not previously tied genetically to ST2, and they confirm the important role of ST2 signaling in asthma.

Melissa H. Bloodworth, Ph.D. ^{a*}

Mark Rusznak ^{b*}

Lisa Bastarache, M.S. ^c

Janey Wang, M.S. ^c

Joshua C. Denny, M.D., M.S. ^c

R. Stokes Peebles, Jr., M.D. ^{a,b}

^aDepartment of Pathology, Microbiology, and Immunology, Vanderbilt University School of Medicine, Nashville, TN, USA

^bDivision of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA

^cDepartment of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA

*These authors contributed equally to this work

References:

1. Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
2. Torgerson, D. G., Ampleford, E.J., Chiu, G.Y., Gauderman, W.J., Gignoux, C.R., Graves, P.E., *et al.* Meta-analysis of Genome-wide Association Studies of Asthma In Ethnically Diverse North American Populations. *Nat Genet.* **43**, 887-892 (2011).
3. Licona-Limón, P., Kim, L.K., Palm, N.W. & Flavell, R.A. T_H2, allergy and group 2 innate lymphoid cells. *Nature Immunology* **14**, 536-542 (2013).
4. Peine, M., Marek, R. M. & Lohning, M. IL-33 in T Cell Differentiation, Function, and Immune Homeostasis. *Trends Immunol.* **37**, 321-333 (2016).
5. Griesenauer, B. & Paczesny, S. The ST2/IL-33 Axis in Immune Cells during Inflammatory Diseases. *Front Immunol.* **8**, [Epub ahead of print] (2017).
6. Grotenboer, N. S., Ketelaar, M.E., Koppelman, G.H. & Nawjin, M.C. *et al.* Decoding asthma: Translating genetic variation in IL33 and IL1RL1 into disease pathophysiology. *J Allergy Clin Immunol.* **131**, 856-865 (2013).
7. Savenije, O. E., Kerkhof, M., Reijmerink, N.E., Brunekreef, B., de Jongste, J.C., Smit, H.A., *et al.* Interleukin-1 receptor-like 1 polymorphisms are associated with serum IL1RL1-a, eosinophils, and asthma in childhood. *J Allergy Clin Immunol.* **127**, 750-756 (2011).
8. UK Biobank: Resources. Available at: <http://www.ukbiobank.ac.uk/resources/>. Accessed Jan 31, 2018.
9. Levescot, A., Flamant, S., Basbous, S., Jacomet, F., Féraud, O., Anne Bourgeois, E., *et al.* BCR-ABL-induced deregulation of the IL-33/ST2 pathway in CD34+ progenitors from chronic myeloid leukemia patients. *Cancer Res.* **74**, 2669-2676 (2014).

Table I. Allergic and asthmatic phenotypes significantly associate with polymorphisms in the ST2 gene.

SNP	MAF (%)	PheWAS Phenotype	PheWAS Phenotype Code	Cases	MAF (%) Cases	Controls	MAF (%) Controls	OR	P value	P value (FDR)
rs1041973	22.41	Allergic conjunctivitis	371.21	420	26.79	25276	22.31	1.27	2.0×10^{-3}	0.50
		Wheezing	512.1	470	19.26	14672	22.66	0.813	1.4×10^{-2}	0.654
		Eosinophilia	288.3	122	16.53	21272	22.40	0.686	3.0×10^{-2}	0.67
		Chronic obstructive asthma with exacerbation	495.11	43	12.79	23477	22.54	0.505	3.4×10^{-2}	0.67
rs4988956	37.78	Asthma	495	1886	34.89	23477	37.92	0.876	2.0×10^{-4}	0.28
		Allergies, other	949	116	27.16	23713	37.68	0.614	1.0×10^{-3}	0.36
rs1420101	38.05	Asthma	495	1886	34.89	23477	37.92	1.20	1.074×10^{-7}	1.5×10^{-4}
		Allergies, other	949	116	27.16	23713	37.68	1.45	5.1×10^{-3}	0.77
		Asthma with exacerbation	495.2	316	42.88	23477	37.82	1.25	6.2×10^{-3}	0.77

Definition of abbreviations: OR= Odds Ratio, MAF= Mean Allele Frequency, FDR=False Discovery Rate Correction, PheWAS= Phenome- Wide Association Study

Table II. Leukemia, acute Myeloid Leukemia, and acute Lymphoid Leukemia significantly associate with polymorphisms in the ST2 gene.

SNP	MAF (%)	PheWAS Phenotype	PheWAS Phenotype Code	Cases	MAF (%) Cases	Controls	MAF (%) Controls	OR	P value	P value (FDR)
rs1420101	38.05	Leukemia	204	1006	40.95	26852	37.92	1.13	5.9×10^{-3}	0.77
		Myeloid Leukemia, acute	204.21	323	41.95	26852	37.92	1.18	3.6×10^{-2}	0.92
rs1041973	22.41	Myeloid Leukemia, acute	204.21	323	26.16	26852	22.23	1.24	1.8×10^{-2}	0.66
		Lymphoid Leukemia, acute	204.11	323	25.85	26852	22.23	1.23	2.4×10^{-2}	0.66
		Leukemia	204	1006	24.20	26852	22.23	1.12	3.7×10^{-2}	0.66

Definition of abbreviations: OR= Odds Ratio, MAF= Mean Allele Frequency, PheWAS= Phenome-wide association study
FDR= False Discovery Rate

Online Repository Methods:

PheWAS Methodology. We used phecodes to define phenotypic features from the EHR.

Phecodes are defined by hierarchical groupings of ICD-9 codes and were developed to enable research based on readily available and computable billing code data for phenome-wide association studies. They have been validated in a broad range of genetic association studies, and have been shown to replicate 66% of tested known associations for a diverse set of phenotypes with an area under the receiver operator characteristic curve of 0.83. The version of phecodes (V1.1) used in this paper, can be downloaded at <http://phewascatalog.org>. We defined cases as individuals with two or more phecodes on unique dates and controls as individuals who do not have a phecode within the control range. For this study, phecode associations with SNPs yielding uncorrected P values lower than 0.05 were considered significant. Further technical details regarding the development and application of PheWAS can be found in our previous publications.^{1,2}

The sequencing was performed using the Illumina human exome genotyping array clustering and quality control protocol developed by the Vanderbilt Epidemiology Center in cooperation with Vanderbilt Technologies for Advanced Genomics Analysis and Research Design. The detailed protocol for this sequencing technique have been previously reported.³

Eosinophilia Phenotype Validation. In this study, eosinophilia refers to peripheral blood eosinophilia as diagnosed and documented by the subject's physician. We performed an analysis of peripheral blood eosinophilia for eosinophilia-associated rs1041973 cases and controls. There was a significant increase in the maximum-recorded percentage of peripheral blood eosinophils in cases (13.5%) compared to controls (3.4%, $p = 3.15 \times 10^{-165}$, Supplemental Table 1).

Replication of Results in Independent Cohorts. To validate the associations that we present in this letter, we repeated our study in three independent cohorts: the UK Biobank, the Vanderbilt BioVU MEGA Project, and the Marshfield Clinic Personalized Medicine Research Project. Of these three, the Vanderbilt BioVU MEGA Project (7,212 subjects) and the Marshfield Clinic Personalized Medicine Research Project (20,000) cohorts were underpowered to detect statistically significant differences. However, we were successful in replicating our allergic and asthmatic phenotype associations with ST2 SNPs in the UK Biobank (500,000 subjects) using self-reporting phenotype codes (Supplemental Table 2). We were unable to confirm our associations to leukemia and lymphoma due to an insufficient number of cases in the cohort.

Online Repository References:

1. Denny, J. C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
2. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–2376 (2014).
3. Guo, Y. *et al.* Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* **9**, 2643–2662 (2014).

Supplemental Table I. Eosinophil percentages for eosinophilia- associated rs1041973 cases and controls used in the study.

Measurement	Cases	Controls	p value
Eos min	3.42	1.60	2.8×10^{-21}
Eos max	13.54	3.69	3.2×10^{-165}
Eos mean	7.04	2.49	1.7×10^{-102}

Definition of Abbreviations: Eos min, max, mean= the minimum, maximum, and mean eosinophil percentages of peripheral blood mononuclear cells.

Supplemental Table II. Allergic and asthmatic phenotype associations with ST2 polymorphisms in the UK Biobank using self reported phenotype codes.

SNP	Self Report Phenotype String	Self Report Phenotype Code	Cases	Controls	Beta	P value
rs10192036	Asthma	6152_8	38791	297991	-1.0x10 ⁻²	8.9x10 ⁻³⁷
	Hay fever, allergic rhinitis, or eczema	6152_9	77891	258891	-1.3x10 ⁻²	9.24x10 ⁻³⁵
rs1041973	Wheeze or whistling in the chest	68531	420	262726	-7.5x10 ⁻³	3.3x10 ⁻¹⁰
rs4988956	Asthma	6152_8	38791	297991	-1.0x10 ⁻²	1.0x10 ⁻³⁶
	Hay fever, allergic rhinitis, or eczema	6152_9	77891	258891	-1.3x10 ⁻²	1.2x10 ⁻³⁴
rs10192157	Asthma	6152_8	38791	297991	-1.0x10 ⁻²	2.2x10 ⁻³⁶
	Hay fever, allergic rhinitis, or eczema	6152_9	77891	258891	-1.3x10 ⁻²	7.0x10 ⁻³⁵
rs10206753	Asthma	6152_8	38791	297991	-1.0x10 ⁻²	8.0x10 ⁻³⁷
	Hay fever, allergic rhinitis, or eczema	6152_9	77891	258891	-1.3x10 ⁻²	9.3x10 ⁻³⁵
rs1420101	Asthma	6152_8	38791	297991	-1.0x10 ⁻²	2.8x10 ⁻³⁸
	Hay fever, allergic rhinitis, or eczema	6152_9	77891	258891	-1.3x10 ⁻²	3.8x10 ⁻²⁷

Definition of Abbreviations: Self Report Phenotype String= Affirmative answers to the question "Has a doctor ever diagnosed you with x?"