RESEARCH NOTE

Two-stage genome-wide association study of chronic rhinosinusitis and disease subphenotypes highlights mucosal immunity contributing to risk

Marcus Soliai, MS[1] (ID), Agnes S. Sundaresan, MD[2], Andréanne Morin, PhD[1], Annemarie G. Hirsch, PhD[2], Catherine Stanhope, MS[1], Jordan Kuiper, PhD[3], Brian S. Schwartz, MD, MS[2,3], Carole Ober, PhD[1] (ID) and Jayant M. Pinto, MD[4]

C hronic rhinosinusitis (CRS) is a prevalent inflammatory disease that results in over $10 billion in healthcare costs annually in the United States. CRS (and its most severe form, nasal polyposis [NP]) overlaps clinically with asthma and aspirin-exacerbated respiratory disease (AERD). Despite having an enormous public health impact, the underlying molecular mechanisms of CRS remain unknown.

That CRS has a genetic component is suggested by familial aggregation and co-occurrence with both monogenic (eg, cystic fibrosis, primary ciliary dyskinesia) and complex (eg, asthma, allergic rhinitis) genetic disorders. Recently, Kristjansson et al reported their large meta-analysis of genome-wide association studies (GWAS) of CRS (n = 5608 cases) and NP (n = 4366 cases) from the United Kingdom and Iceland.[1] Cases and non-CRS/non-NP controls (n > 700,000) were defined by *International Classification of Disease—tenth revision* (ICD-10) codes. Ten genome-wide significant loci for NP and 2 for CRS were identified. The finding of only 2 genome-wide significant loci despite ∼10,000 cases suggests that substantial clinical heterogeneity may have masked the effects of other risk loci, and that phenotyping using diagnosis codes likely captures a range of conditions with overlapping symptoms. Thus, alternative approaches are needed.

Using an epidemiologic framework, we targeted patients receiving primary care at Geisinger, a health-care provider to ∼1.6 million people in Pennsylvania and New Jersey. Survey information from the Chronic Rhinosinusitis Integrative Studies Program (CRISP)[2] and associated electronic health records (EHR) identified 483 European American (EA) adults who reported at least 2 of 6 cardinal symptoms of CRS most or all of the time for at least 3 months, where at least 1 of the symptoms was obstruction, anterior nasal discharge, or posterior discharge (epidemiologic definition of CRS)[3] and who met at least 1 of the following criteria: (i) reported surgery for CRS or NP in any 1 of 4 CRISP surveys (3 individuals); (ii) evidence of previous sinus surgery on sinus computed tomography (CT) scan (127 individuals); or (iii) a Lund-Mackay score of ≥3 on sinus CT scan (91 individuals) (see Supplemental Information online). For controls, we selected 2057 EAs without CRS from the MyCode Community Health Initiative,[4] a Geisinger system-wide biobank, based on absence of NP,

[1]*Department of Human Genetics, University of Chicago, Chicago, IL;* [2]*Department of Population Health Sciences, Geisinger, Danville, PA;* [3]*Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD;* [4]*Section of Otolaryngology–Head and Neck Surgery, Department of Surgery, University of Chicago, Chicago, IL*

Correspondence to: Jayant M. Pinto, MD, Section of Otolaryngology–Head and Neck Surgery, Department of Surgery, University of Chicago, Room E103, MC 1035, 5841 South Maryland Avenue, Chicago, IL 60637; e-mail: jpinto@surgery.bsd.uchicago.edu

A

B

Chromosome	rsID	Position*	Nearby genes†	Allele‡	AF	OR	P value	eGene	QTL Tissue
2p13.3	rs3833578	2:69088930	LINC01888, BMP10	TA/T	0.061	0.3733	3.22×10^{-06}	-	-
4p16.2	rs6818775	4:4648918	STX18-AS1, STX18, LOC01928279	A/G	0.161	0.4854	1.68×10^{-07}	-	-
7q36.3	rs3935208	7:157843523	PTPRN2, LOC100506585, MIR595	C/A	0.547	1.680	1.58×10^{-07}	-	-
9q31.1	rs10820254	9:105488979	LINC00587, CYLC2	G/A	0.303	0.6084	5.21×10^{-06}	CYCL2	Testis
12q24.23	rs11068795	12:118495743	WSB2, RFC5, VSIG10	C/A	0.238	1.737	2.15×10^{-06}	WSB2	Skeletal Muscle, Skin
21q21.3	rs2830160	21:27651254	MARCKSP1, CYYR1-AS1	G/A	0.140	0.5224	4.96×10^{-06}	AP000230.1	Brain, Adrenal Gland

* SNP position according to Genome Reference Consortium Build 37 (hg19)
† The gene in which the SNP is located is indicated first followed by the genes located upstream and downstream, respectively; for intergenic SNPs, only the upstream and downstream genes are included.
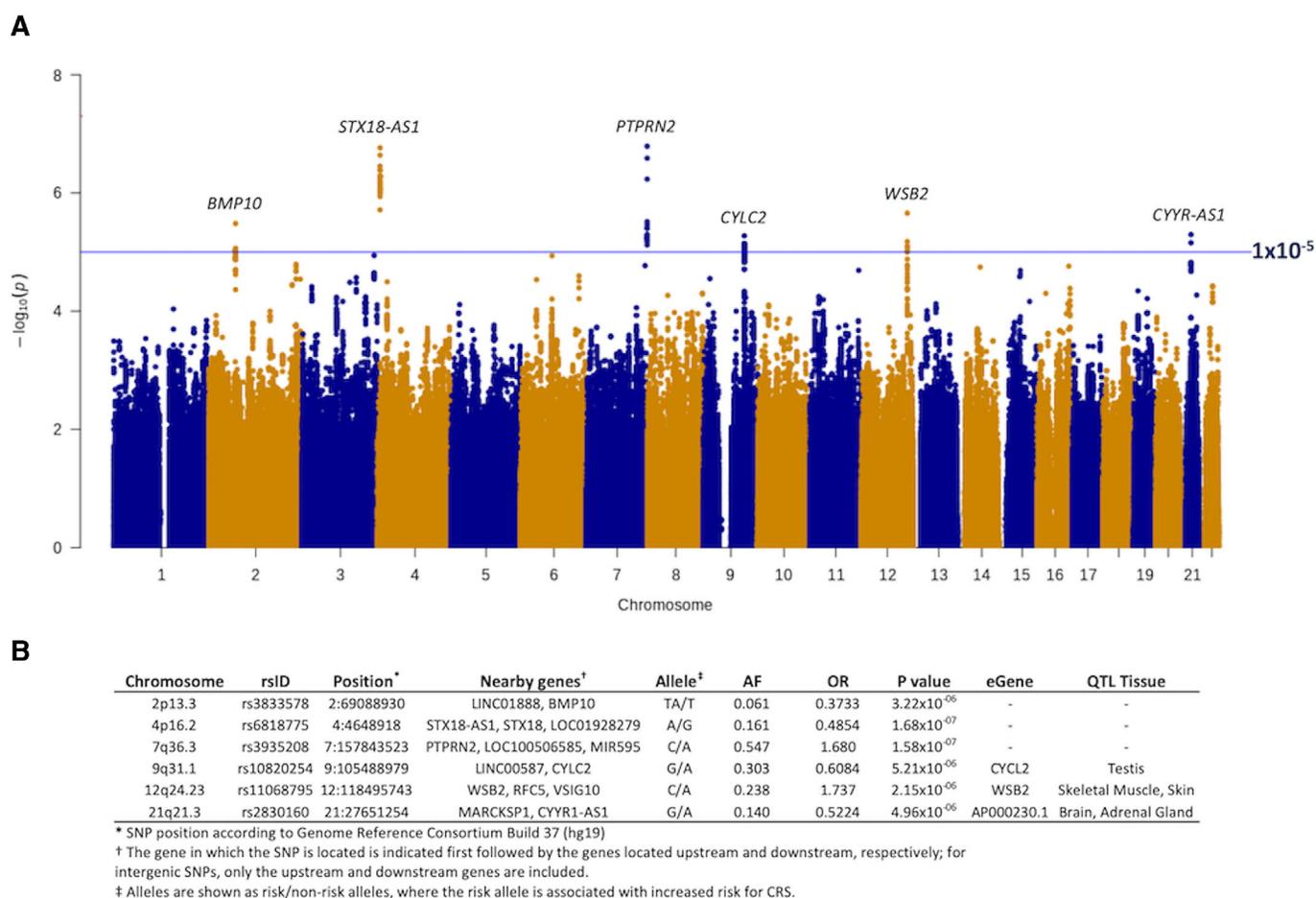‡ Alleles are shown as risk/non-risk alleles, where the risk allele is associated with increased risk for CRS.

FIGURE 1. CRS GWAS results. (A) Manhattan plot showing chromosome positions along the x axis and –log(p value) on the y axis. The horizontal blue line indicates suggestive significance ($p < 1 \times 10^{-5}$). The closest gene to the sentinel SNP at each locus reaching suggestive significance is shown. (B) Description of the 6 sentinel SNPs in regions reaching suggestive significance. CRS = chronic rhinosinusitus; GWAS = genome-wide association studies; SNP = single-nucleotide polymorphism.

CRS, asthma, or allergic rhinitis diagnoses and no report of a sinus CT scan, nasal endoscopy, or sinonasal surgery in the EHR (see Table S1 in Supplemental Information). Genotypes were obtained on all subjects (Illumina bead-chip arrays) and imputed to yield >4.5 million single-nucleotide polymorphisms (SNPs), and then used to perform this first GWAS of CRS in the United States (see Supplemental Information). This study was approved by the institutional review boards of the University of Chicago and Geisinger; all patients provided informed consent.

To maximize both sample size and the detailed phenotyping available in some subjects, we performed a 2-stage analysis. In the first stage, we included all cases in a CRS GWAS. This revealed 82 suggestive significant SNPs ($p < 1 \times 10^{-5}$) at 6 loci (Fig. 1A). The sentinel SNP at each locus is described in Figure 1B. An SNP in linkage disequilibrium (LD; $r^2 > 0.9$; GRB, 1000 Genomes Project) with 1 of the 2 SNPs associated with CRS and NP from the Kristjansson et al[1] study was not associated with CRS in our GWAS ($p > 0.90$); the second SNP associated with CRS and NP in that study was not imputed or in LD with SNPs in our study.

Because defining CRS by medical records may be imprecise, we defined CRS subphenotypes in 646 CRISP subjects with sinus CT scans, including 172 cases in the GWAS, applying latent class analysis (LCA) to create subgroups based on patterns of sinus opacification,[5] to identify clinical and immunopathogenic subgroups of CRS.[6] We identified 3 latent classes, which we referred to as "no/mild opacification," "localized opacification," and "diffuse opacification" (see Table S2). In the second stage, we performed association studies between the 6 sentinel GWAS SNPs (Fig. 1B) and the latent classes (Fig. 2), considering the no/mild opacification group as the reference. One SNP at the 12q24.23 locus (rs11068795) was associated with the localized opacification class phenotype ($p_{corrected} = 0.002$) but not with the diffuse opacification class phenotype ($p_{corrected} = 1.0$), after Bonferroni correction for 6 tests. Another SNP at the 9q31.1 locus (rs10820254) was associated with the diffuse ($p_{corrected} = 0.048$) but not the localized opacification ($p_{corrected} = 1.0$) class phenotype. Thus, defining phenotypes based on imaging of inflammatory characteristics in the sinuses highlighted and

A

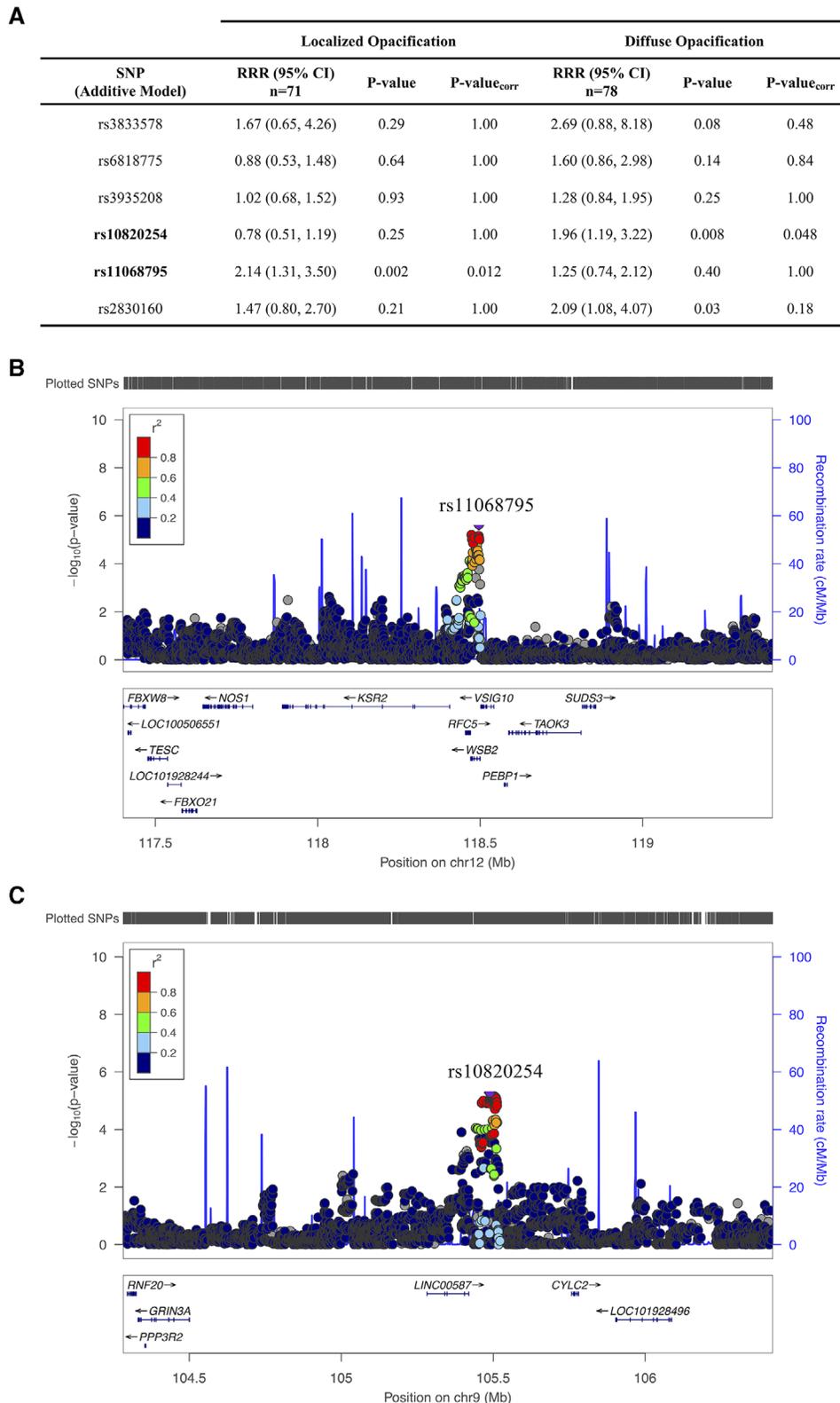| SNP (Additive Model)	Localized Opacification			Diffuse Opacification		
	RRR (95% CI) n=71	P-value	P-value$_{corr}$	RRR (95% CI) n=78	P-value	P-value$_{corr}$
rs3833578	1.67 (0.65, 4.26)	0.29	1.00	2.69 (0.88, 8.18)	0.08	0.48
rs6818775	0.88 (0.53, 1.48)	0.64	1.00	1.60 (0.86, 2.98)	0.14	0.84
rs3935208	1.02 (0.68, 1.52)	0.93	1.00	1.28 (0.84, 1.95)	0.25	1.00
rs10820254	0.78 (0.51, 1.19)	0.25	1.00	1.96 (1.19, 3.22)	0.008	0.048
rs11068795	2.14 (1.31, 3.50)	0.002	0.012	1.25 (0.74, 2.12)	0.40	1.00
rs2830160	1.47 (0.80, 2.70)	0.21	1.00	2.09 (1.08, 4.07)	0.03	0.18

B

C

FIGURE 2. Associations of sentinel SNPs with LCA-defined subtypes. (A) Associations between 2 LCA-defined subtypes based on computed tomography scans. Each subtype is compared with the "no/mild opacification" group (n = 191). The *p* values are corrected for 6 tests. Rows show the sentinel SNPs at each suggestive significant GWAS locus, and its association with each of the subtypes. Results of multinomial logistic regression models, including surgical status, sex, and physician-diagnosed patient-reported allergic rhinitis as covariates, are shown for an additive model. Significant SNPs and results are shown in bold. The LocusZoom plots for the 2 significant SNPs in (A) are shown in (B) (rs11068795) and (C) (rs10820254). The SNPs are shown in a ±1-Mb window with chromosome and gene positions shown on the x axis and the −log$_{10}$ of the *p* value and the recombination rate (CMMb) are shown on the left and right y axis, respectively. SNPs are colored according to the estimated linkage disequilibrium (*r*^{2}) relative to each lead SNP (purple diamond) and calculated using the 1000 Genomes European panel. CI = confidence interval; LCA = latent class analysis; RRR = relative risk ratio.

validated 2 loci that were associated with CRS in the first-stage GWAS, despite the significantly reduced sample size.

The SNP associated with the localized opacification class phenotype, rs11068795, is associated with the expression of the *WSB2* gene in skin ($p = 2.3 \times 10^{-7}$) and skeletal muscle ($p = 7.1 \times 10^{-17}$).[7] WSB2 functions in the proteasomal degradation of target proteins as a negative regulator of the interleukin-21 receptor (IL-21R), a molecule associated with microbe-related epithelial conditions (eg, inflammatory bowel disease and *Helicobacter pylori* infection), which are themselves associated with CRS. IL-21R expression is also upregulated in atopic dermatitis patients,[8] who manifest *Staphyloccocus aureus* colonization barrier dysfunction. This is consistent with previous work showing upregulation of IL-21 in CRS with NP.[9]

The SNP associated with the diffuse opacification class phenotype, rs10820254, is an intergenic SNP located between the *CYLC2* gene and a gene encoding a long intergenic non–protein coding RNA, *LINC00587*. *CYLC2* is a sperm-specific cyclin 2 and LINC00587 is expressed most highly in the testis.[7] Thus, rs10820254 may contribute to CRS pathobiology through long-range interactions with the promoters of distal genes, which remain unidentified to date.

Our study was limited by sample size, which was small by GWAS standards. As a result, we likely missed associations with minor effects. To address this, we performed a 2-stage analysis to identify new susceptibility loci. Moreover, the initial self-reported CRS diagnoses may be prone to error. Any resulting misclassification could have further limited power. However, self-reported, physician-diagnosed disease is used commonly in GWAS of other conditions, including asthma.[10] Last, we could not look separately at episodic vs stable CRS due to sample size constraints, although there is no evidence to suggest that these CRS phenotypes have different genetic etiologies.

In conclusion, using a 2-stage population approach, we identified variation at 2 loci potentially involved in CRS, including a connection to mucosal immunity, supporting a role for the upper airway surface as a key site in disease development. Last, inclusion of subjects with careful phenotyping in genetic studies of CRS may generate insights into underlying mechanisms and identify potential therapeutic targets.

References

1. Kristjansson RP, Benonisdottir S, Davidsson OB, et al. A loss-of-function variant in ALOX15 protects against nasal polyps and chronic rhinosinusitis. *Nat Genet*. 2019;51:267-276.

2. Hirsch AG, Stewart WF, Sundaresan AS, et al. Nasal and sinus symptoms and chronic rhinosinusitis in a population-based sample. *Allergy*. 2017;72:274-281.

3. Fokkens WJ, Lund VJ, Mullol J, et al. EPOS 2012: European Position Paper on Rhinosinusitis and Nasal Polyps 2012. A summary for otorhinolaryngologists. *Rhinology*. 2012;50:1-12.

4. Carey DJ, Fetterolf SN, Davis FD, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med*. 2016;18:906-913.

5. Muthén LK, Muthén BO. Mplus User's Guide: Statistical Analysis with Latent Variables. 7th ed. Los Angeles, CA: Muthén & Muthén; 1998-2012.

6. Meng Y, Lou H, Wang C, Zhang L. Predictive significance of computed tomography in eosinophilic chronic rhinosinusitis with nasal polyps. *Int Forum Allergy Rhinol*. 2016;6:812-819.

7. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648-660.

8. Jin H, Oyoshi MK, Le Y, et al. IL-21R is essential for epicutaneous sensitization and allergic skin inflammation in humans and mice. *J Clin Invest*. 2009;119:47-60.

9. Xiao L, Wei Y, Zhang YN, et al. Increased IL-21 expression in chronic rhinosinusitis with nasal polyps. *Clin Exp Allergy*. 2015;45:404-413.

10. Demenais F, Margaritte-Jeannin P, Barnes KC, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet*. 2018;50:42-53.