

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

Marcus Soliai, MS¹ , Agnes S. Sundaresan, MD², Andréanne Morin, PhD¹, Annemarie G. Hirsch, PhD², Catherine Stanhope, MS¹, Jordan Kuiper, PhD³, Brian S. Schwartz, MD, MS^{2,3}, Carole Ober, PhD¹  and Jayant M. Pinto, MD⁴

© 2020 The Authors. International Forum of Allergy & Rhinology published by Wiley Periodicals LLC on behalf of American Academy of Otolaryngic Allergy and American Rhinologic Society.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Key Words:

Chronic Rhinosinusitis; Rhinosinusitis; Statistics

How to Cite this Article:

Soliai M; Sundaresan AS; Morin A; Hirsch AG; Stanhope C; Kuiper J; Schwartz BS; Ober C; Pinto JM. Two-stage genome-wide association study of chronic rhinosinusitis and disease subphenotypes highlights mucosal immunity contributing to risk. *Int Forum Allergy Rhinol.* 2021;00:1-4.

Chronic rhinosinusitis (CRS) is a prevalent inflammatory disease that results in over \$10 billion in health-care 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,

¹Department of Human Genetics, University of Chicago, Chicago, IL;

²Department of Population Health Sciences, Geisinger, Danville, PA;

³Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; ⁴Section 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

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Funding source for the study: National Institute of Allergy and Infectious Diseases (U19 AI106683).

Potential conflict of interest: None provided.

Received: 13 July 2020; Revised: 19 October 2020; Accepted: 19 October 2020

DOI: 10.1002/alr.22731

View this article online at wileyonlinelibrary.com.

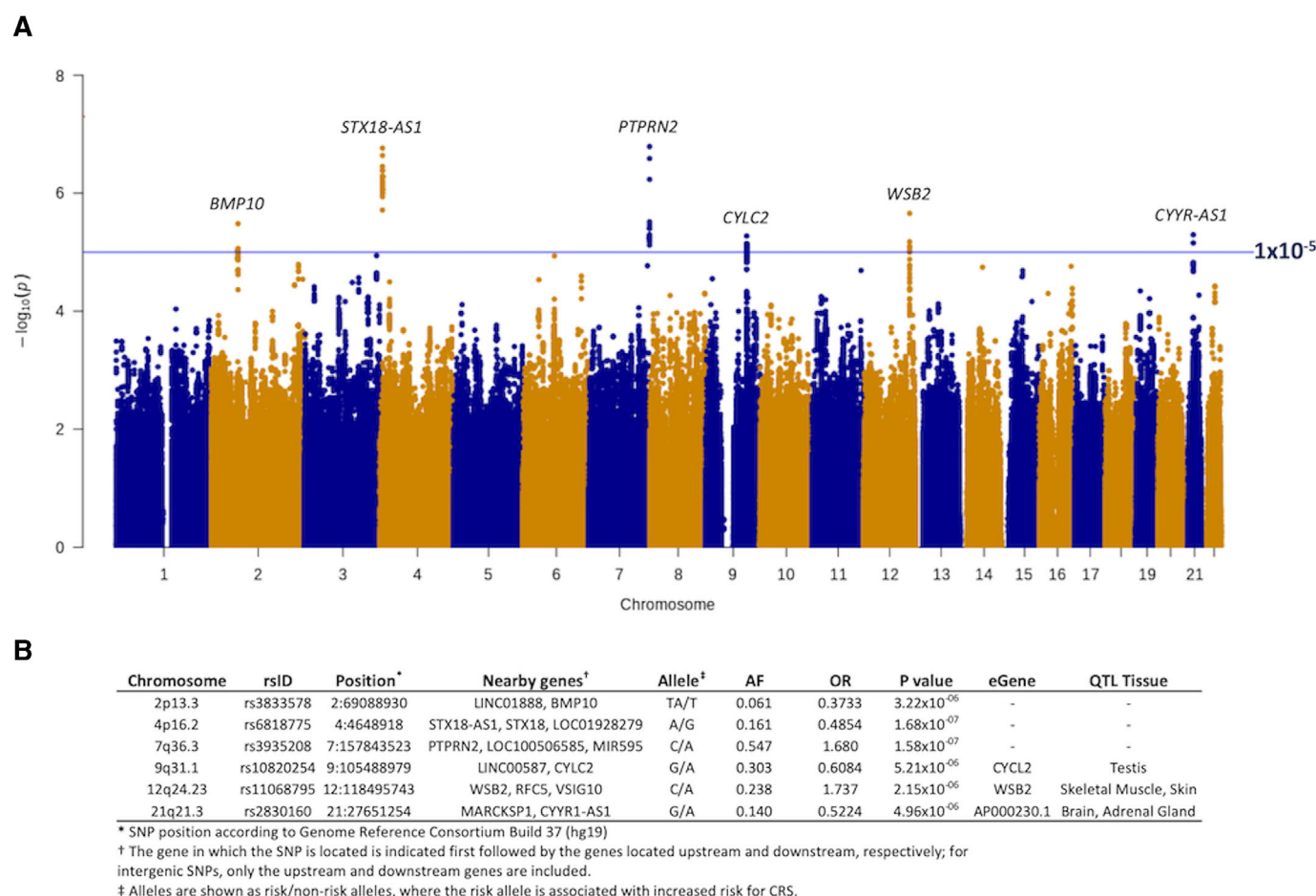


FIGURE 1. CRS GWAS results. (A) Manhattan plot showing chromosome positions along the x axis and $-\log(p \text{ value})$ on the y axis. The horizontal blue line indicates suggestive significance ($p < 1 \times 10^{-5}$). The closest gene to the sentinel SNP at each locus reaching suggestive significance is shown. (B) Description of the 6 sentinel SNPs in regions reaching suggestive significance. CRS = chronic rhinosinusitis; 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 al¹ 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 ($p_{\text{corrected}} = 0.002$) but not with the diffuse opacification class phenotype ($p_{\text{corrected}} = 1.0$), after Bonferroni correction for 6 tests. Another SNP at the 9q31.1 locus (rs10820254) was associated with the diffuse ($p_{\text{corrected}} = 0.048$) but not the localized opacification ($p_{\text{corrected}} = 1.0$) class phenotype. Thus, defining phenotypes based on imaging of inflammatory characteristics in the sinuses highlighted and

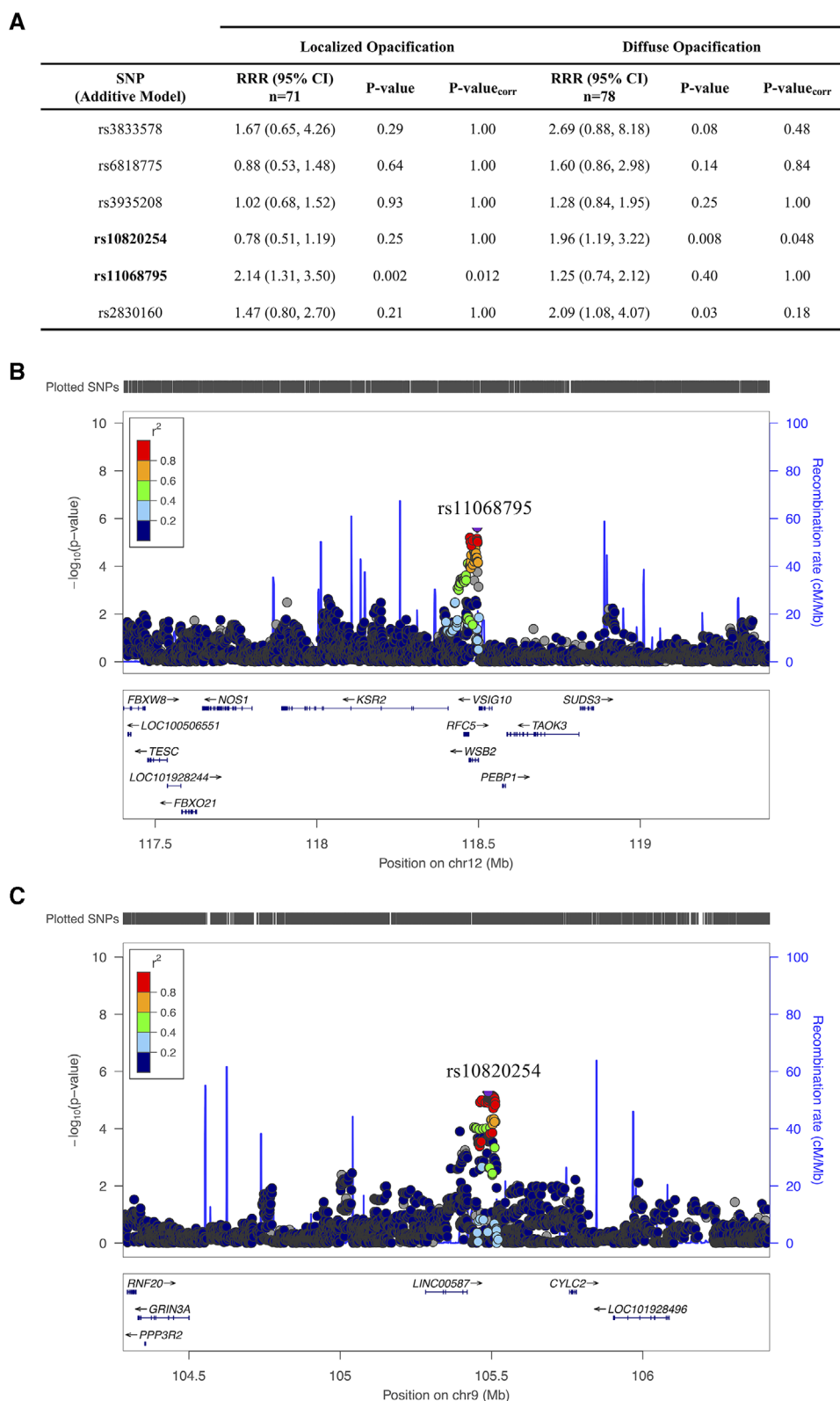


FIGURE 2. Associations of sentinel SNPs with LCA-defined subtypes. (A) Associations between 2 LCA-defined subtypes based on computed tomography scans. Each subtype is compared with the “no/mild opacification” group ($n = 191$). The p values are corrected for 6 tests. Rows show the sentinel SNPs at each 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 (cM/Mb) 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 *Staphylococcus 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

- 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.
- 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.
- 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.
- 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.
- Muthén LK, Muthén BO. Mplus User's Guide: Statistical Analysis with Latent Variables. 7th ed. Los Angeles, CA: Muthén & Muthén; 1998-2012.
- 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.
- Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348:648-660.
- 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.
- 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.
- Demenais F, Margeritte-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.

Supplementary Information

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

Marcus Soliai, Agnes S. Sundaresan, Andreeanne Morin, Annemarie G. Hirsch,
Catherine Stanhope, Jordan Kuiper, Brian S. Schwartz, Carole Ober, Jayant M. Pinto

Table of Contents

<i>Methods</i>	2
Geisinger population and databases.....	3
Genotyping and Imputation.....	5
Genome-Wide Association Studies.....	6
Latent Class Analysis on Sinus CT.....	7
Statistical analysis for the second-stage.....	7
<i>Tables</i>	9
S1. Study Sample Characteristics.....	10
S2. Demographic information for GWAS and LCA Samples	11
<i>Figures</i>	12
S1. Geisinger population and databases.....	13
<i>References</i>	14

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

Methods

Geisinger population and databases

Subjects were selected from 200,769 adult primary care patients from Geisinger, a health system that serves 45 counties in Pennsylvania. The Geisinger primary care population is representative of the general population in the region. Subjects were selected from this patient pool based on diagnostic codes in the electronic health record (EHR) for chronic rhinosinusitis (CRS) and asthma/allergy, and questionnaires were sent to a random sample of 23,700 people, enriched for ethnic minorities and those with certain diagnostic and therapeutic codes.¹ The 7,847 who returned the questionnaire constituted the Chronic Rhinosinusitis Integrative Study Program's (CRISP) study cohort. Rich longitudinal EHR data, four seasonal follow-up questionnaires are available for these subjects in addition to a questionnaire at the time of CT scan. DNA from blood and saliva for a subset of samples in the Geisinger biobank, a MyCode Community Health Initiative, are available for these subjects. CRS was defined by questionnaire responses based on the European Position Paper on Rhinosinusitis (EPOS) epidemiologic criteria²: nasal discharge (anterior or posterior) or nasal obstruction, most or all of the time for at least three months, plus at least one other symptom of that frequency and duration (nasal discharge, nasal obstruction, smell loss, facial pain, or facial pressure). Among the CRISP cohort, 1,866 had current CRS (as defined by EPOS in the prior three months) and 2,093 had past CRS (as defined by EPOS in their lifetime prior to three months). Of these individuals with CRS (ever): (1) 1,499 had a MyCode consent and a DNA sample in the Geisinger biobank; opt-out letters were mailed to these individuals prior to using their sample. (2) 372 had MyCode consents but lacked a DNA sample; these individuals were mailed a consent for the CRISP genetics study with post cards requesting a saliva sample. DNA was collected from 126 of the 372 individuals via saliva collection kits that were mailed to participants (79 samples) or from

whole blood (47 samples) that were collected in the interim. (3) Individuals that signed-up for MyCode were sent out mailing in batches, prioritizing those with current CRS, participation in the CT study, or had longitudinal data, until the study sample size was attained. 204 individuals returned consents and 81 were prioritized and mailed DNA kits, of whom 56 returned the kit. Overall, 1,661 of the 1,681 individuals with CRS (ever) had high-quality DNA and were genotyped at the University of Chicago. This subset of the CRISP cohort (CRS by EPOS criteria + genotypes) were the source population for the CRS GWAS cases. The GWAS cases were those with CRS as defined by EPOS epidemiologic criteria of European American ancestry, and who had objective evidence of disease on sinus CT determined by an otorhinolaryngologist (ENT; 127 individuals) or had self-reported sinus surgery for CRS or nasal polyposis (NP) in any one of four CRISP surveys (at baseline, in the six- or 16-month surveys or the CT survey; 3 individuals), had evidence of prior sinus surgery on a sinus CT or had Lund-Mackay (LM) score of ≥ 3 on sinus CT (91 individuals).

Controls for the GWAS were derived from the MyCode Community Health Initiative³, which includes a Geisinger-system-wide biobank of participants who were genotyped as part of the DiscovEHR cohort, a Geisinger-Regeneron collaboration. Controls were selected from this larger cohort based on the following criteria: European American ancestry and absence of (i) an ICD-9 code for NP, CRS, asthma, or allergic rhinitis, and (ii) a Current Procedural Terminology (CPT) code for sinus CT scan, endoscopy, or surgery in the EHR. Twenty-five of these controls were also part of the CRISP cohort who never had CRS. These samples are described in

Supplementary Table 1.

Among the individuals who completed the baseline CRISP questionnaire (described above), 3,269 subjects were invited for sinus CT scans by stratified random sampling to enrich

for those with sinus symptoms.⁴ The strata were based on the presence of CRS symptoms as defined by EPOS (CRS)² and latent class analysis (LCA) categories on symptoms reported in the questionnaire.⁵ Of these subjects, 646 completed the sinus CT scans and LCA was performed again on their sinus CT findings, and subjects were classified into three CT-defined subgroups based on patterns of observed sinus opacification: no/mild opacification LCA class, localized LCA class, and diffuse LCA class. Of the 646 LCA-classified subjects, 341 also had genetic information. Of the 341 with both sinus CT and genetic information, 339 had current or past CRS based on symptom criteria (similar to the case definition in the GWAS). Genetic association studies using the sentinel SNP at each of six suggestive significant loci in the GWAS were performed on the 339 subjects assigned to the latent classes. The LCA cases were those with CRS by epidemiologic definition AND localized LCA class (n = 71) or diffuse LCA class (n = 78), while the LCA controls had CRS by epidemiologic definition but no/mild opacification LCA class (n = 191). In order to differentiate subjects in the two stages of analysis with different sets of cases and controls, the first stage of GWAS is described using the terms GWAS cases and controls, and the second stage of the study using LCA cases and controls; 51% of those in the LCA genetics analysis were also GWAS cases (**Supplementary Figure 1**).

Genotyping and Imputation

The GWAS cases and LCA cases and controls were genotyped at the University of Chicago using the Human Core Exome Array (550,224 SNPs). The controls were genotyped by the Geisinger Biobank using either the Illumina Human Core Exome (533,456 SNPs) or the Human Omni Express Arrays (731,306 SNPs). Quality control (QC) was performed for cases and controls separately for each of the genotyping platforms (excluding SNPs with HWE <

0.00001, call rate < 0.95, and MAF = 0). Samples from each of the three platforms were then merged and QC checks were applied in the pooled sample, as was previously performed for each of the genotyping platforms. After this QC, 222,149 overlapping markers for 3,718 individuals were available for analysis. Ancestral principal components analysis (PCA) was performed using 649 ancestral informative markers that overlapped with HapMap release 3 and were included on all of the genotyping arrays.

Phasing and imputation were performed in the merged dataset using the ShapeIt2⁶ and Impute2⁷ software packages, respectively. Variants were imputed in 5 Mb windows across the genome against the 1000 Genomes Phase 3 haplotypes (Build 37; October 2014). Individuals were grouped as European based on how they related to the HapMap reference panel and the k-means clustering of ancestral PCs, using the kmeans⁸ function in R, to confirm that all subjects were of European ancestry. After imputation, additional QC steps were performed with gtool,⁹ including the exclusion of X and Y chromosome-linked SNPs and variants with info score < 0.8, MAF < 0.05, missingness > 0.05 and a probability score < 0.9. Probability scores were converted to dosages for 10,436,943 SNPs used in downstream analyses.

Genome-Wide Association Studies

A GWAS was performed using the Genome-wide Efficient Mixed Model Association (GEMMA) software toolkit,¹⁰ fitting a multivariate linear mixed model¹¹ to test for marker association with CRS while controlling for population stratification. A standardized relatedness matrix was estimated from the genotypes using 0.02 missingness and 0.05 MAF thresholds. A multivariate linear mixed model was applied using age, sex, ancestry (ancestry PCs 1-3), and relatedness (standardized relatedness matrix) as covariates, and 0.02 missingness and 0.05 MAF

thresholds for 10,436,943 SNPs. In total, 4,528,849 SNPs were analyzed in 483 CRS cases and 2,057 non-CRS controls of European descent (**Supplementary Table 1**). P-values from the likelihood ratio test are reported. The most significant SNP at each of six suggestive significant loci from this GWAS was referred to as the sentinel SNP, and used in the second-stage LCA association study.

Latent class analysis on sinus CT

LCA was used to create unobservable subgroups based on patterns of observed variables in the sinus CT ENT reads. From the LM score on the sinus CT reported by ENT, binary indicators were created (0 and ≥ 1) and included in a series of LCA models fit with an increasing number of classes. Appropriateness of model fit was assessed using multiple fit statistics. The three-class model had the best fit including the “no/mild opacification,” “localized opacification,” and “diffuse opacification” classes. The LCA analysis was performed using Mplus v.8.1 (Muthén & Muthén, Los Angeles, CA). Traditional LM score is in a linear scale and there are no universally accepted cut-offs for clustering. However, LCA identified subgroups in the sinus CT analysis that likely have clinical significance, making it a good phenotype for this study (**Supplementary Table 2**).

Statistical analysis for the second-stage

The aim of the LCA genetic analysis was to study the associations of the sentinel SNPs at GWAS loci with surgical CRS and CRS evidence on CT or objective evidence of CRS (CRSo) with the outcome of LCA CT classes. We used chi-square tests to compare the proportion of individuals with different categories of LCA CT by sex, race/ethnicity, smoking, Medical

Assistance, health insurance status (a surrogate for family socioeconomic status), subject reported physician-diagnosed allergic rhinitis and asthma, atopy, migraine headache,¹ anxiety sensitivity index (ASI) (below median vs. at or above median), and history of sinus surgery. Analysis of variance was used to examine associations of CRS CT subgroups with continuous variables like age and LM score. Base model-building was performed by adding the above predictors individually to the model and retaining those with significant associations. The final base model included surgical status, sex and physician-diagnosed subject-reported allergic rhinitis. Multinomial logistic regression was used for the associations of genetic data and LCA CT classes, adjusted for base model variables in an additive model with p-values corrected for six tests. All analyses were performed using Stata 15.1 (StataCorp LLC, Texas, USA).

187

188

189

190

191

192

193

194

195

Tables

Table 1. Demographic and clinical information for GWAS and LCA subjects.

Characteristic	GWAS source dataset (n = 1661)	GWAS cases (n = 483)	GWAS controls (n = 2057)	No/mild opacification LCA (n = 190)	Localized opacification LCA (n = 71)	Diffuse opacification LCA (n = 78)
Age in years as of 4-2014, mean (SD)	55.40 (14.02)	56.63 (13.51)	57.35 (16.68)	55.93 (12.05)	58.37 (12.91)	58.46 (12.98)
Male, n (column %)	531 (31.97)	204 (42.24)	462 (22.46)	39 (20.53)	26 (36.62)	40 (51.28)
*Self-reported race, n (column %)						
White	1622 (97.65)	477 (98.76)	2054 (99.85)	183 (96.32)	71 (100.00)	76 (97.44)
Hispanic	39 (2.35)	6 (1.24)	3 (0.15)	7 (3.68)	0	2 (2.56)
Smoking, n (column %)						
Current	248 (14.93)	59 (12.22)	313 (15.22)	16 (8.42)	6 (8.45)	8 (10.26)
Former	513 (30.89)	142 (29.40)	662 (32.18)	51 (26.84)	25 (35.21)	31 (39.74)
Never	900 (54.18)	282 (58.39)	1082 (52.60)	123 (64.74)	40 (56.34)	39 (50.00)
BMI kg/m ² , mean (SD)	31.48 (7.60)	31.45 (7.11)	31.07 (7.50)	32.08 (7.95)	31.04 (6.84)	32.18 (6.60)

*All individuals that were included in the GWAS clustered with Europeans according to the k-means clustering of the ancestral principal components and were therefore combined for the analysis.

Abbreviations: BMI = Body Mass Index; GWAS = Genome-Wide Association Study; LCA = Latent Class Analysis; SD = Standard Deviation

Table 2. Clinical profile of the LCA subjects.

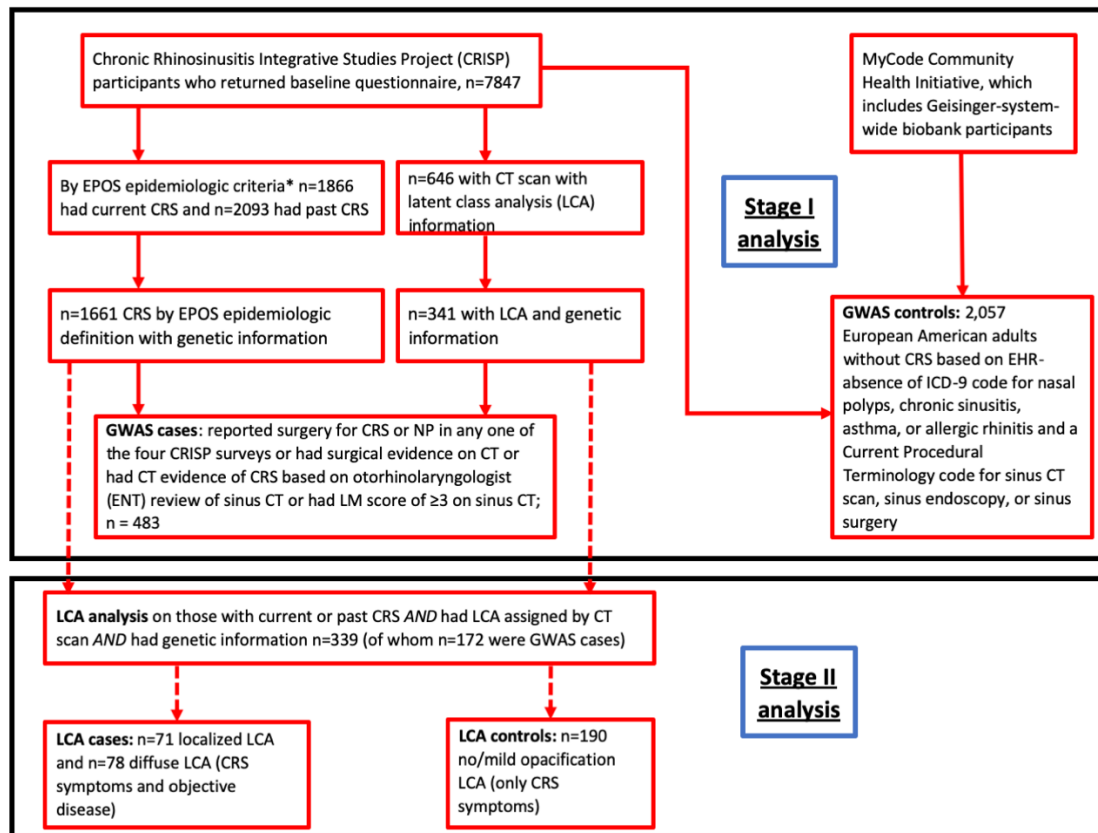
Characteristic	No/mild opacification (n =190)	Localized opacification (n= 71)	Diffuse opacification (n= 78)	P-values* for LCA
Age in years, mean (SD) at time of CT	58.20 (12.04)	60.57 (12.91)	60.65 (12.92)	0.21
Males, n (column %)	39 (20.53)	26 (36.62)	40 (51.28)	<0.001
White, n (column %)	183 (96.32)	71 (100.00)	76 (97.44)	0.26
Smoking, n (column %)				
Current	16 (8.42)	6 (8.45)	8 (10.26)	0.22
Former	51 (26.84)	25 (35.21)	31 (39.74)	
Never	123 (64.74)	40 (56.34)	39 (50.00)	
Medical assistance, n (column %)	22 (11.58)	7 (9.86)	9 (11.54)	0.92
Asthma self-reported physician diagnosis, n (column %)	65 (34.21)	20 (28.17)	32 (41.03)	0.26
Allergic rhinitis- self-reported physician diagnosis, n (column %)	98 (51.58)	46 (64.79)	50 (64.10)	0.06
Atopy (self-reported skin prick positivity), n (column %)	82 (43.16)	35 (49.30)	29 (37.18)	0.33
Migraine, n (column %)	78 (41.05)	23 (32.39)	20 (25.64)	0.046
Anxiety sensitivity index ASI score (below median vs. at or above), n (column %)	76 (48.72)	25 (44.64)	36 (55.38)	0.48
Original LM score, mean (SD)	0.22 (0.53)	1.75 (1.09)	7.14 (3.41)	<0.001
Surgery on CT, n (column %)	29 (15.26)	11 (15.49)	34 (43.59)	<0.001

*p-values < 0.05 indicates at least one latent class significantly differs from the others with respect to the selected characteristic.

Abbreviations: CT = Computed Tomography; LCA = Latent Class Analysis; LM = Lund-Mackay; SD = Standard Deviation

Figures

Supplementary Figure 1. Flowchart of CRS case and non-CRS control subject inclusion in analyses performed in this study.



*European Position Paper on Rhinosinusitis (EPOS) epidemiological definition for CRS: At least two of six symptoms (obstruction, post-nasal drip, posterior nasal drainage, facial pain, facial pressure, smell loss) most or all of the time for at least three months, where at least one of the symptoms was obstruction, anterior nasal discharge, or posterior discharge

Abbreviations: CRS= chronic rhinosinusitis; CT = computed topography; EHR = electronic health record; GWAS = genome-wide association study; ICD = International Classification of Diseases; LM= Lund-Mackay; NP= Nasal polyposis

References

1. Hirsch AG, Stewart WF, Sundaresan AS, Young AJ, Kennedy TL, Scott Greene J, et al. Nasal and sinus symptoms and chronic rhinosinusitis in a population-based sample. *Allergy* 2017; 72:274-81.
2. Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, et al. EPOS 2012: European position paper on rhinosinusitis and nasal polyps 2012. A summary for otorhinolaryngologists. *Rhinology* 2012; 50:1-12.
3. Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med* 2016; 18:906-13.
4. Hirsch AG, Nordberg C, Bandeen-Roche K, Tan BK, Schleimer RP, Kern RC, et al. Radiologic sinus inflammation and symptoms of chronic rhinosinusitis in a population-based sample. *Allergy* 2019.
5. Sundaresan AS, Hirsch AG, Young AJ, Pollak J, Tan BK, Schleimer RP, et al. Longitudinal Evaluation of Chronic Rhinosinusitis Symptoms in a Population-Based Sample. *J Allergy Clin Immunol Pract* 2018; 6:1327-35 e3.
6. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013; 10:5-6.
7. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* 2011; 1:457-70.
8. Hartigan JAAW, M. A. Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics* 1979:100-8.
9. Freeman C, Marchini J. GTOOL: A program for transforming sets of genotype data for use with the programs SNPTEST and IMPUTE, Oxford, UK. GTOOL: A program for transforming sets of genotype data for use with the programs SNPTEST and IMPUTE, Oxford, UK.
10. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012; 44:821-4.
11. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods* 2014; 11:407-9.