Inflammatory and infectious upper respiratory diseases associate with 59 genomic loci that link to type 2 inflammation genes

Elmo Saarentaus
University of Helsinki https://orcid.org/0000-0002-8475-7187
Juha Karjalainen
The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
Joel Rämö
Helsinki Institute of Life Science
Tuomo Kiiskinen
Institute for Molecular Medicine Finland https://orcid.org/0000-0002-6306-8227
Aki Havulinna
Helsinki Institute of Life Science
Juha Mehtonen
Helsinki Institute of Life Science
Heidi Hautakangas
Helsinki Institute of Life Science
Sanni Ruotsalainen
Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki
https://orcid.org/0000-0002-1631-4069
Sanna Salmi
Helsinki Institute of Life Science
Matti Pirinen
University of Helsinki https://orcid.org/0000-0002-1664-1350
Mitja Kurki
Helsinki Institute of Life Science
Samuli Ripatti
Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki
Mark Daly
Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki
Tuula Palotie
Helsinki Institute of Life Science
Antti Mäkitie
Helsinki Institute of Life Science

Aarno Palotie (aarno.palotie@fimm.fi)
Helsinki Institute of Life Science

Genetics Article

Keywords: upper respiratory diseases, autoimmune disease, inflammation genes

Posted Date: July 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-654053/v1

Inflammatory and infectious upper respiratory diseases associate with 59 genomic loci that link to type 2 inflammation genes

Authors: Elmo C. Saarentaus[1], Juha Karjalainen[1,2,3], Joel T. Rämö[1], Tuomo Kiiskinen[1,4], Aki S. Havulinna[1,4], Juha Mehtonen[1], Heidi Hautakangas[1], Sanni Ruotsalainen[1], FINNGEN, Sanna Toppila-Salmi[5,] Matti Pirinen[1,6,7], Mitja Kurki[1,2,3], Samuli Ripatti[1,2,3], Mark Daly[1,2,3], Tuula Palotie[8,9], Antti Mäkitie[10], Aarno Palotie[1,2,11]*

Affiliations:

[1]Institute for Molecular Medicine Finland FIMM, HiLIFE, University of Helsinki, Helsinki, Finland.

[2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

[3]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.

[4]Finnish Institute for Health and Welfare, Helsinki, Finland.

[5]Skin and Allergy Hospital, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

[6]Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

[7]Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[8]Orthodontics, Department of Oral and Maxillofacial Diseases, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland;

[9]Oral and Maxillofacial Diseases, Helsinki University Hospital, Helsinki, Finland

[10]Department of Otorhinolaryngology – Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

[11]Analytic and Translational Genetics Unit, Department of Medicine, Department of Neurology and Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA.

*Corresponding author. Email: aarno.palotie@fimm.fi

One Sentence Summary: Inflammatory upper respiratory diseases share genetic risk loci and correlate genetically with each other and with diseases of the immune system

Abstract:

Inflammatory and infectious upper respiratory diseases (IURDs; ICD-10: J30-J39), such as diseases of sinonasal tract, pharynx and larynx, are growing health problems yet their genomic similarity is not known. We analyzed genome-wide association to eight IURDs (61,195 cases) among 260,405 FinnGen participants, meta-analyzing diseases in four groups based on an underlying genetic correlation structure. We aimed to understand which genetic loci contribute to susceptibility to IURDs in general and its subtypes. We detected 59 independent genome-wide significant (GWS) loci, distinguishing impact on sinonasal or pharyngeal diseases, or both. Fine-mapping implicated non-synonymous variants in 16 genes, including 10 linked to immune-related diseases. Phenome-wide analysis implicated asthma and atopic dermatitis at sinonasal disease loci and inflammatory bowel diseases, and other immune-mediated disorders at pharyngeal disease loci. IURDs also genetically correlated with autoimmune diseases such as rheumatoid arthritis, autoimmune hypothyroidism, and psoriasis. Finally, we associated separate gene pathways in sinonasal and pharyngeal diseases that both contribute to type 2 immunological reaction. We show shared heritability among IURDs that extends to several immune-mediated diseases with diverse mechanisms, such as type 2 high inflammation.

Main Text:

INTRODUCTION

Inflammatory and infectious upper respiratory diseases (IURD) affect the sinonasal tract, pharynx, and larynx, and include diseases such as chronic tonsillitis, allergic rhinitis, and chronic rhinosinusitis. They lead to increased morbidity[1,2] and costs[3], and to the highest public health burden in the world[4] by serving as the main route of infection to the body, and by their connection to non-communicable diseases, such as asthma[5-7], autoimmune diseases[8,9], cardiovascular diseases[10] and obesity[11]. Genetic predisposition[12-15] together with environmental megatrends such as the COVID-19 pandemic[16-18], Western lifestyle[19], urbanization[20,21], global warming[22] and dysbiosis[23,24] influence the burden of IURDs. IURDs often co-exist[25-28], and they have shown overlapping mechanisms[5,6,29,30]. Understanding the genetic (dis)similarities behind IURDs can remarkably improve preventive actions and therapies, and reduce the burden of IURDs and related diseases[2,31].

IURDs are characterized by an etiology related to recurrent infections and dysbiosis[20,21,24] leading to chronic and treatment-resistant diseases[32] with acute and even life-threatening exacerbations. IURDs involve inflammation in the nasal cavity, such as vasomotor and allergic rhinitis (VAR), both characterized by hyperresponsiveness to stimuli[33]; non-specific chronic rhinitis, nasopharyngitis and pharyngitis (CRNP) and nasal septal deviation (NSD); and in the adjoining paranasal sinuses, such as chronic rhinosinusitis (CRS) with or without nasal polyps (NP)[6]. Allergic rhinitis (AR) is a part of an allergic disease entity involving allergic asthma, atopic dermatitis, allergic conjunctivitis and food allergy[27,34,35]. IURDs also encompass other

diseases of the pharynx such as chronic laryngitis and laryngotracheitis (CLT), chronic diseases of tonsils and adenoids (CDTA), and peritonsillar abscess (PA). Previous genetic studies of IURDs and related immune responses have largely focused on rare variants[36] and the *HLA* region[37,38]. IURD-related GWAS have been reported of CRS and NP[39], tonsillectomy and childhood ear infections[40,41], cold sores, mononucleosis, strep throat, pneumonia and myringotomy[40], hay fever[39-42], and of infective diseases caused by specific airway-related microbes such as pneumococcus[43] and staphylococcus aureus[44]. However, no prior research has analyzed shared genetic contributions of IURDs.

The FinnGen study is a large biobank study including both genetic and lifelong health record data from all participants, thus allowing the investigation of potentially shared and distinct genetic landscape associated to IURDs. This provides an opportunity both for GWASs as well as for cross-disease analyses to better understand potential shared genetic contributors. We aimed to study genetic predispositions to recurrent, chronic and complicated IURDs. We hypothesized that, on one hand, shared genetic variants contribute to IURD susceptibility in general, and some variants contribute more to distinct IURD phenotypes. To test this hypothesis, we analyzed genome-wide association of IURD cases in the FinnGen study (release 6/Aug 2020), a nation-wide collection of genotyped samples from Finnish individuals. Our study sample included 260,405 individuals of all ages, where we focused on cases of specialist-diagnosed IURDs (n = 61,197), including their more specific diagnosis. We tested the genetic associations across IURDs to highlight shared and distinct genetic contributions among IURDs. Finally, we compared the genome-wide association of IURDs and phenotypes to other anatomically related

and systemic immunological disorders (such as chronic periodontitis) linked with the same genetic loci.

RESULTS

Genome-wide association of inflammatory and infectious upper respiratory diseases

We performed genome-wide association of all IURD cases (n = 61,197, ranging from 2,623 to 29,135 per phenotype) in FinnGen. We genotyped and imputed 16,355,289 single-nucleotide genetic variants in 260,405 Finnish individuals of all ages. We used a logistic mixed model with the SAIGE software[45] (see Methods) to detect genome-wide association between 61,197 cases of IURDs (Table 1, Supplementary Figure 1) and 199,208 controls, and set as covariates age, genetic sex, principal components (PCs) 1–10, and genotyping batch. This resulted in 698 genome-wide significant (GWS, p<5e-8) SNPs in 20 loci (Table 2). Fine-mapping using the SuSiE software[46] on the 19 non-*HLA* loci identified 21 independent credible sets of causal variants with at least one GWS SNP.

Next, we analyzed genome-wide association in IURD phenotypes for each ICD-10 category and used the summary statistics to study the genetic correlation. Inspired by the IURD GWAS results, we used LD Score regression[47] to investigate the underlying genetic structure (Figure 1A). A high genetic correlation (r_g > 75 %) distinguished two clusters: I) VAR, CRS, NP, and NSD ($r_g \geq 78$ %); II) CDTA and PA (r_g = 79 %). Using a threshold of r_g > 90 % further distinguished a genetically linked subgroup of known comorbid disorders[26]: III) VAR, CRS, NP. We denoted these IURD groups as “sinonasal diseases” (I), “pharyngeal diseases” (II) and “chronic inflammatory sinonasal diseases” [CISDs, (III), Figure 2]. All subgroup and phenotype analyses used the same set of individuals without a disease in the IURD category (n = 199,208) as controls.

We then performed a GWAS analysis in each of the IURD groups (Figure 2) using the same pipeline as above. Sinonasal diseases (n = 25,235) associated with 13 GWS loci (Supplementary Table 1), repeating seven of the loci associated with all IURDs and identified six additional GWS loci (Supplementary Figure 2A). GWAS of CISDs (n = 19,901) identified 15 GWS loci (Supplementary Table 2), repeating seven loci associated with all IURDs and four loci associated with sinonasal diseases (the parent group), and four additional GWS loci (Supplementary Figure 2B). GWAS of pharyngeal diseases (n = 33,157) identified 25 GWS loci (Supplementary Table 3), repeating 10 loci associated with all IURDs and 15 additional associations (Supplementary Figure 2C). Four IURD-associated loci (near *CCDC54*, *PZP*, *FBXO33* and *KRT19*) were not GWS in the subgroup GWASs. Finally, we analyzed the genome-wide association of specific IURD phenotypes (VAR, CRNP, CRS, NP, NSD, CDTA, PA, and CLT), detecting an additional 13 GWS associations (Supplementary Figure 3). In total, genome-wide association of IURD detected 59 independent GWS loci (Figure 1B, Supplementary Table 4).

To highlight the similar impacts of many detected loci among IURDs, we used hierarchical clustering of lead SNP z-scores to group specific loci (Figure 1B). The clusters grouped largely into sinonasal and pharyngeal diseases, further corroborating the divide into sinonasal diseases and pharyngeal diseases. This also distinguished shared sinonasal and pharyngeal impact on four loci (1q21.3, 5p13.2, 9q33.3, 17q12). In total, 22 of 58 non-*HLA* loci had a co-directional association with at least one other IURD phenotype.

In addition to the main phenotypes linked to the upper respiratory tract, we also analyzed genome-wide association to two oral inflammatory diseases that have been associated[48-50] with

IURDs: diseases of pulp and periapical tissues (DPPT; ICD-10 K04) and to chronic periodontitis (CP; ICD-10 K05.3). We observed two GWS loci for DPPT (48,687 cases vs 211,718 controls), one locus implicating *HORMAD2* with a credible set overlapping that of CDTA at the same locus (Supplementary Table 5). One GWS SNP (rs80193913) was detected for CP (14,631 cases vs 245,774). GWAS of DPPT subphenotypes repeated the two loci in pulpitis (K04.0, n = 18,139) and necrosis of pulp (K04.1, n = 10,168).

For replication, we compared our phenotypes to those available at the public PanUK database[51]. We mapped the phenotypes VAR, CRS, NP, and CDTA to corresponding UKB endpoints (Methods). These four phenotypes included 43 non-*HLA* loci-phenotype associations in the FinnGen analysis. Forty-one associations had a GWS SNP with available summary statistics in PanUK (Supplementary Table 6). 35 of these 41 FinnGen associations had a similar direction of effect in the PanUK analysis. We found that the effect size differed ($p < 0.05$) only for the CRS association to the locus 5q22.1 (near *WDR36*) and the NP association to the locus 1q21.3 (near *ARNT*). Lead SNPs of 17 loci, including all 12 NP loci, showed significant association ($p < 0.05$) in the PanUK analysis, and all 17 of these were codirectional with the observed effect in Finngen.

Phenome-wide analysis (PheWAS) of lead SNPs

The high degree of genetic correlation observed in GWASs was also seen in phenome-wide analysis (PheWAS) of lead SNPs (Supplementary Table 7). Lead variants of the 58 non-*HLA* GWS loci were associated with 2,861 endpoints in FinnGen. The Bonferroni-corrected level of phenome-wide significance (PWS) was $p < 1.748e\text{-}5$. In line with earlier epidemiological

studies, the PheWAS-implicated disorders of the immune system for loci of sinonasal and pharyngeal disease. There was PWS association to three endpoints of infectious and inflammatory disorders of the upper respiratory tract that were not included in our definition of IURD: acute upper respiratory tract infections, non-suppurative otitis media, and eustachian salpingitis and obstruction. These endpoints associated with loci near *IL18R1*, *LTBR*, and *TNFRSF13B*, respectively. Asthma was associated with 14 loci associated with NP, CRS, and VAR – similar to what has been reported at previously identified loci associated with CRS with NP (CRSwNP)[39]. Pharyngeal disease loci near genes *IL7R* and *DYSF* also associated with asthma. Phenome-wide association to autoimmune diseases was observed near *RAB5B*, *HORMAD2* and *NEK6*, and to inflammatory bowel disease near *ARNT*, *CDC42SE2*, and *GATA3*. The CDTA-associated locus near *ABO* was associated with venous thromboembolism, duodenal ulcer, diseases of the gall bladder, diabetes-related endpoints, and diverticular disease of the intestine, among a total of 53 PWS disease endpoints. The CDTA- and PA-associated locus *LTBR* also associated with increased risk for acute appendicitis [OR = 1.07 (1.04–1.10)] and reduced risk for nonsuppurative otitis media [OR = 0.90 (0.85–0.95)].

GWAS identifies non-synonymous coding variants in 16 genes

The fine-mapped credible sets included VEP-predicted non-synonymous variants in 16 protein-coding genes (Table 3), including nine genes with previous immunomodulatory function. Four of these variants are low frequency variants enriched in the Finnish population. The *IL1RL1* and *ZPBP2* missense variants have been previously associated with Type 2 high childhood asthma[52] and adult-onset asthma[53], respectively. The Finnish-enriched rs144651842 variant is in the well-

known asthma locus *IL4R*[54,55]. The *SLC22A4* and *FUT2* variants are linked with Crohn's disease[56,57].

Functional variants also mapped to five known immune deficiency genes. The CDTA-associated 17p11.2 locus, previously also linked with tonsillectomy[40], identifies the non-synonymous variant rs72553883-T. This Finnish-enriched missense variant in the gene *TNFRSF13B* (encoding the protein TACI)[58,59] is linked to common variable immune deficiency (CVID) (variant MIM no 604907.0002) and primary antibody deficiency[60]. The missense variant rs117556162 in immune deficiency-linked *CARMIL2* (phenotype MIM no. 618131) decreases risk for CDTA, as does the missense variant rs2272676 in CVID-linked *NFKB1*[61]. A missense variant in exon 6 of the severe combined immunodeficiency-linked[62] *IL7R* (phenotype MIM no. 608971) decreases risk for IURDs [OR = 0.96 (0.94–0.98)]. In total, IURD-associated non-synonymous variants in five genes – *NFKB1*, *IL7R*, *CARMIL2*, *TNFRSF13B*, and *FOXN1* – are included in the IUIS list of Mendelian immune disorder genes[15].

In addition to the immunological effects of non-synonymous variants, the Finnish-enriched missense variant in gene *G6PC* as well as the variant in the pleiotropic *FUT2* gene increase risk for statin medication in FinnGen – the latter variant also for type 1 diabetes. In participants where the *ZNF417* indel rs774674736-A was directly genotyped and not imputed (n = 197,006), there were no homozygotic carriers detected ($p_{HWE} = 0.44$).

Characterization of loci

We used the Functional Mapping and Annotation pipeline (FUMA)[63] to further annotate predicted non-intronic and non-intergenic SNPs and eQTLs using the UKB v2 Europeans LD

structure. The FUMA analysis associated non-intronic SNPs in 157 protein-coding genes, including six of the missense variants in Table 3. This implicated non-synonymous variants in seven additional genes, and additional non-synonymous variants in *IL1RL1* and *FUT2* (Supplementary Table 8). We then compared SNPs with expression data from GTEx v8[64] and DICE[65], which showed change in the expression of a total of 264 genes in all 58 non-*HLA* loci in total. Differing expression of 27 genes (14 protein-coding) in immunological tissues was associated with SNPs in 10 pharyngeal disease loci, and expression of C19orf73 in CD4+ T_{H17} cells was associated with rs281379 in the CRS-associated 19q13.33 locus. The expression levels differed in two additional genes linked with Mendelian immune disorders: *IKBKB* expression in skeletal muscle was associated with 38 independent SNPs associated with CDTA; and *STAT2* expression in the tibial nerve was associated with rs705702 and three other NP-associated SNPs. Full results from FUMA positional and eQTL analysis are reported in Supplementary Table 9.

We tested gene enrichment with MAGMA[66] using summary statistics for all IURD phenotypes (Methods). We detected 96 gene-phenotype associations (Supplementary Table 10) for 74 genes, after correcting for multiple testing. MAGMA detected significant variant enrichment within 44 genes that were within the 58 non-*HLA* GWS loci. Eleven genes were enriched in more than one IURD phenotype. Two genes, *WDR36* and *TSLP*, were associated with VAR, CRS and NP through the shared locus on 5q22.1. Thirteen associated genes were not close to any of the GWS loci, including *IL2RB* (linked with NP and CRS), *ST5* and *ESR1* (both linked with CDTA), and a cluster of four genes at 20q13.33 associated with VAR.

We next tested for enrichment of genes in 4,761 curated gene sets and 5,917 GO terms. Enrichment of MAGMA-identified genes highlighted gene sets involved with immune function (Table 4), including major histocompatibility class II receptor activity, and regulation and production of interleukins 4 and 13. The tumor necrosis factor 2 pathway, which spans 16 recognized genes, includes 10 genes associating with CDTA. The recognized associations encompass variants in genes *TRAF2*, *TRAF3*, *TANK*, *TNFRSF1B,* and *RIPK1*, all involved in producing the intracellular components of TNF receptor 2.

Bayesian analysis of shared variant effects

We established the loci-specific shared and distinct genetic impact by comparing the associations in phenotype-specific GWAS using a Bayesian framework (see Methods and Supplementary data) for lead SNPs. Briefly, this framework tests the probability of hypothesized association models for a variant using summary statistics of the GWASs being compared, taking into account the overlapping cases and controls between phenotypes[67]. The framework allowed us to evaluate the following models: the null model, where the variant explains no part of any of the phenotype variation; the fixed model, where the variant has one fixed effect that is the same for all phenotypes; the correlated model, where the variant has a correlating effect on all phenotypes; and models where the impact is to one phenotype only.

The Bayesian framework distinguished a subdivision for the detected loci in most cases, providing evidence that some of the variant associations were more disease-specific than others. For the 19 non-*HLA* lead SNPs detected in IURD for example (Figure 3A-B), a shared effect was supported (P(Fixed or Correlated) > 75 %) for five of the 19 non-*HLA* loci. Six loci were likely

only impacting pharyngeal diseases, two only sinonasal diseases, and four likely both tonsil and sinonasal diseases. Thus 9/19 loci were considered shared between sinonasal and pharyngeal diseases, with possible effect on CL and CRNP from five loci – the remaining loci likely being more specific in their impact. Two loci remained uncertain: rs11406102 had a less clear general impact (P(Fixed or Correlated) = 73.3 %), and rs1837253 impacted sinonasal diseases with an uncertain effect on other phenotypes. In a similar vein for sinonasal diseases (Figure 3C-D), consisting of CISD and NSD, all tested models supported an impact on sinus diseases for all variants, and possible impact on NSD for three lead SNPs as well. The pharyngeal disease lead SNPs showed a consistent impact for 20 SNPs, with two SNPs being likely CDTA-specific and three SNPs impacting CDTA and possibly PA (Figure 4A-B). Strikingly, all CISD lead SNPs were either consistent or highly correlated in their effect among the three subphenotypes VAR, CRS and NP (Figure 4C-D).

Genetic correlation with other implicated phenotypes

Genetic correlation analysis beyond IURD highlighted genome-wide links with susceptibility to infection, asthma and allergic diseases. We analyzed genetic correlation using LD Score regression[47], estimating correlating impacts between IURD phenotypes and diseases associated in PheWAS analysis (Supplementary Figure 4). Sinonasal diseases in particular formed a cluster of genetic correlation with asthma, allergic conjunctivitis and atopic dermatitis (Supplementary Figure 4A). The recurring links to autoimmune diseases in several loci translated to genetic correlation with rheumatoid arthritis, mainly with DPPT [r_g = 58.6 % (95 % CI 27.7–89.5 %); p = 0.00020] and CRS [r_g = 57.6 % (27.4–87.1 %); p = 0.00020] (Supplementary Figure 4B). Other tested diseases, such as non-suppurative otitis media and sleep apnea, largely clustered

separately despite significant correlations to specific IURD phenotypes (Supplementary Figure 4C). Finally, when comparing to PheWAS-linked inflammatory intestinal diseases, CRS and CDTA showed genetic correlation with diverticular disease and appendicitis (Supplementary Figure 4D).

Since our investigation focused on host susceptibility to infection, we also compared our results to the summary statistics of the COVID-19 host genetics initiative[18], noting two shared loci at *NFKBIZ* and *ABO*. We investigated the genetic correlation of IURDs and their associated FinnGen endpoints, based largely on pre-pandemic diagnoses, with three COVID-19 endpoints. COVID-19 hospitalization (B2) in particular linked with chronic periodontitis [r_g = 57.2 % (10.7–100.0); p = 0.016], DPPT [r_g = 41.8 % (12.5–71.0); p = 0.0051], all pneumonias [r_g = 34.9 % (3.2–66.7); p = 0.031], CRNP [r_g = 42.6 % (11.7–73.4); p = 0.0068] and hospital discharge record of unspecified acute upper respiratory infections [r_g = 55.8 % (16.3–95.3); p = 0.0055] (Supplementary Figure 5). Pharyngeal or sinonasal diseases were not significantly associated.

DISCUSSION

To understand the genetic predisposition landscape of infectious and inflammatory upper respiratory diseases, we genome-wide analyzed these diseases both in groups and individually. In total, we identified 59 loci, of which 23 have not been previously reported to associate with any of the IURD diseases. Among the 59 loci, our fine-mapped credible sets identified 16 coding variants, including four that are enriched in the Finnish population. In addition, the FUMA pipeline identified 13 additional coding variants using a UK-based LD reference. Cross-disease analyses combined chronic diseases of the sinonasal, oral and pharyngeal regions. We showed that genetic structure distinguished sinonasal diseases and pharyngeal diseases, with a partly overlapping genetic background. Our study also includes the first GWASs of chronic diseases of tonsils and adenoids, peritonsillar abscess, and diseases of pulp and periapical tissues.

Our findings indicate overlapping genetic etiologies that extend beyond the previously reported genetic links between chronic rhinosinusitis and nasal polyps[39] to tonsillar endpoints as well. We observe both locus-specific and genome-wide associations between sinonasal, oral and pharyngeal inflammatory and infectious conditions. The associations implicate immunological pathways and links to immune-mediated diseases beyond the confines of the upper respiratory system.

Of the 59 observed genome-wide significant loci, only thirteen loci were uniquely observed in a single IURD phenotype GWAS. Out of the remaining 46 loci, 41 had highly similar odds ratios in two or more phenotypes, even if the signal did not reach the genome-wide significance threshold in all diseases. This is in line with previous epidemiological and histopathological

evidence that highlights links between inflammatory sinonasal diseases[5,6,26,42]. Similarly, pharyngeal diseases associate with 13 loci previously linked to tonsillectomy[40,41] as well as two loci associated with self-reported strep throat and childhood ear infections[40]. While there is a shared genetic contribution for five loci to all IURDs, the underlying genetic structure distinguishes between sinonasal diseases and pharyngeal diseases. This is supported by several lines of evidence, from genome-wide correlation to loci-specific z-score based hierarchies and Bayesian meta-analysis. In addition to the known links among chronic inflammatory sinonasal diseases, we observe a shared heritability between the clinically distinct chronic (CDTA) and acute (PA) pharyngeal diseases, including sixteen loci with shared impact.

Overall, genetic correlation analysis illustrated the genetic landscape linking IURDs with asthma, reflecting the co-existence of IURDs and asthma; about half of AR, CRS and NP patients have asthma[6]. We detected shared genetic risk for NP and CRS, which is in line with previous observations[39]. In subjects with NP, gene set enrichment associated pathways related to regulation and production of interleukins 4 and 13, which are hallmarks of Type 2 inflammation[68] and have shown to be associated with asthma[69] and to be functionally relevant in CRSwNP patients[70]. *Ex vivo* cultured nasal basal cells have been shown to retain intrinsic Type-2 high memory of IL-4/IL-13 exposure, which could be decoded via clinical blockade of the IL-4 receptor α-subunit *in vivo*[70]. We demonstrated a genetic landscape linking IURDs (such as NP, CRS and AR) with asthma and allergic diseases[39,71]. We found that chronic inflammatory sinonasal diseases associate with the 17q21 locus (*GSDMP*) as well as *TSLP*, *IL33* and its receptor, *IL1RL*, which all have previously shown to be associated with asthma[72,73], and have also shown to be functionally relevant in asthma models[74-77].

Pharyngeal diseases implicate many genes linked with immune deficiency. Non-synonymous variants were implicated in 20 loci, highlighting nine genes linked with immune deficiency and immune-mediated disorders. Interestingly, in some genes (*NFKB1*, *CARMIL2*, and *IL7R*) with previously established risk variants for immune deficiency[15], we identify non-synonymous variants with decreased risk for IURDs. Differing expression linked two additional immune deficiency-linked genes (*IKBKB* with CDTA and *STAT2* with NP), thus linking eleven immunomodulatory genes to IURDs in total.

Beyond the above-described trends, there were also associations with autoimmune disorders. Diseases such as rheumatoid arthritis correlated genome-wide with CRS and DPPT, highlighting the multitude of immunopathological mechanisms with manifestations in the upper respiratory tract. Immune-mediated disorders, such as asthma and inflammatory bowel diseases, also replicated in phenome-wide context. Among specific pathways, we implicate the tumor necrosis factor 2 pathway as a viable target for further study in the analysis of chronic diseases of tonsils and adenoids. This furthers the findings of Tian et al[40], who previously identified genetic links between tonsillectomy and the intestinal immune network for IgA production. We also replicate a shared locus near *HORMAD2* between CDTA and type 1 diabetes, and extend IBD-associated[78] CDTA loci to *SLC45A1* and *PIM3*. Beyond specific loci, we reported enrichment of CDTA-associated variants in 10 out of 16 genes involved with the TNFR2 pathway, and many intracellular genes of the canonical NFκB pathway[79]. Notably, the TNFR2 pathway has been found to modulate allergic inflammation[80].

We observed shared impact with other infectious disorders. Immune deficiency was also evident in loci-specific impact on infectious disorders, specifically non-suppurative otitis media and appendicitis. These two infections also genetically correlated with CRS and CDTA. Two of the novel loci described herein – the *ABO* cluster that associates with CDTA and the PA locus closest to the gene *NFKBIZ* – have been implicated with COVID-19 severity[18,81]. PheWAS analysis of the *ABO* gene cluster shows wide-ranging phenotypic implications (Supplementary Table 7), in line with its well described pleiotropic effects[82]. The fine-mapped set of CDTA-linked variants near *ABO* include rs923383567-C [a.k.a. rs657152-C, linked with COVID-19 severity[81]] and rs879055593-C, the latter of which was linked with interleukin-4 driven pathogenesis in a recent multitrait analysis[83]. The *NFKBIZ* locus has two SNPs with near-equal posterior inclusion probability in fine-mapping: rs1456200 and rs1456202 (Supplementary Figure 6). The cryptic LD structure suggests that further work is needed to fine-map this region. While pharyngeal diseases showed little general genetic correlation with COVID-19 in this analysis (Supplementary Figure 5), it is interesting to note the genetic correlation with oral infections and CRNP, although only few loci could be identified in these disorders. The implications of these results require further study and replication.

This study has some limitations. Firstly, the FinnGen study cohort is collected based on legacy samples variably representing certain aspects of the population, with new participants being recruited mainly in the tertiary health care setting. For these reasons, the study cohort is not a true population sample and thus comorbidity analyses should not be interpreted from an epidemiological angle. Second, the IURD phenotypes are diagnosed by specialists, often in hospital setting, and thus likely quite accurate but are therefore subject to ascertainment bias

(collider bias) with other disorders – a feature of study design that can inflate correlation estimates with other diseases. Thirdly, the ICD-10-based disease endpoints used here differ somewhat from current clinical practice (e.g. non-allergic rhinitis and AR, CRSsNP and CRSwNP). However, as the genetic association to disease biology does not necessarily follow clinical manifestations, and there is notable previous success using this approach[39], we therefore find these categories appropriate to highlight the genetic similarities and differences. Finally, while the NP analysis in the PanUK was well-powered for replication, the effective sample sizes for the other three diseases (VAR, CRS, and CDTA) were not sufficient for reliable replication analysis of many of the lead SNPs of these diseases.

Using lifelong national register data, we identified 59 loci associated with different upper respiratory diseases. These loci identified genes involved in immunological (such as Type 2) mechanisms and immune-mediated diseases. We observed both shared and distinct genetic contribution among different chronic inflammatory upper respiratory diseases, between IURDs and other systemic immune-mediated disorders, and between IURDs and two oral inflammatory diseases, providing genetic insight to earlier clinical and epidemiological observations.

MATERIALS AND METHODS

Study design

The FinnGen study is an on-going nationwide collection of Finnish genetic samples, combining genome information with digital health care and registry data. Participants include legacy samples from previous studies recruited for on-going research, maintained by the Biobank of the Finnish Institute for Health and Welfare (THL), and recent biobank samples recruited at university hospitals across Finland. In the present study, we included samples from 271,341 participants released in August 2020. Registry data used here included disease diagnoses and performed operations from the Care Register for Health Care (THL), the Primary Health Care Register (THL), the Causes of Death Register (Statistics Finland) and the Drug Reimbursement Register (KELA, the Social Insurance Institution of Finland).

Participants in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and the start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol Nr HUS/990/2017. The FinnGen study approval permits and biobank access decisions are listed in Supplementary table 11.

Genotype and Sample QC

Samples were genotyped using called for a total of 271,341 individuals. In total, 12 different type of DNA chips were used to analyze participants in 78 batches. In genotyping, we removed SNPs with high missingness (>2%), minor allele count less than 3 and Hardy-Weinberg equilibrium (p_{HWE} < 1e-6). We removed samples with non-Finnish heritage in PC analysis, duplicated/twins, incomplete phenotypic information or mismatch between reported and imputed genetic sex. The final post-QC sample count was 260,405 (147,061 females and 113,344 males). Genotypes were imputed based on a Finnish reference panel developed in-house (Palta et al, manuscript in preparation), using whole-genome sequencing data from 3,775 Finns in the SISu[84] reference panel.

Genome-wide association analyses

We used the SAIGE software[45] for running mixed model logistic regression genome-wide on 16,355,289 variants. We used age, sex, the first 10 PCs and genotyping batch as covariates. We analyzed genome-wide association of cases of IURD (any, n = 61,197) against 199,208 controls in the FinnGen dataset. Also, when analyzing the eight individual IURD phenotypes, we excluded all IURD cases from controls, so the same controls (n = 199,208) were used in all IURD phenotype analyses. These IURD phenotypes did not include J38 (“Diseases of vocal cords and larynx, not elsewhere classified”) and J39 (”Other diseases of upper respiratory tract”), which were included in the larger IURD category but were not separately analyzed.

Characterization of Loci

After the initial detection of genome-wide associated SNPs, we chose lead SNPs based on lowest p-value, and GWS SNPs in the same locus were grouped based on genomic distance < 2Mb, r^2 > 0.1 with lead SNP. We used SuSiE[46] for detection of credible sets of causal variants, with a Finnish-based reference panel [Sequencing Initiative Suomi[84]] for LD structure and imputation. Only credible sets with at least one genome-wide significant (GWS; p < 5e-8) SNP were considered.

We used the FUnctional Mapping and Annotation pipeline (FUMA)[63] to characterize GWS loci using ANNOVAR for functional annotation, selecting for non-intronic and non-intergenic annotations. The UK Biobank release 2b (10k European) was used for the reference panel population. FUMA successfully mapped 38 of the 43 non-HLA GWS loci. We mapped eQTLs for GWS SNPs in all recognized loci using GTEx v8 [all tissues excluding testis[64]] and DICE[65] databases in the FUMA pipeline. All genes (including non-coding) were included in the functional and eQTL analyses. Due to the loss of annotation of some Finnish-specific variants, we performed a manual search of GWS SNPs in loci not tagged by FUMA.

Phenome-Wide Association Study (PheWAS)

We performed a phenome-wide association study on each of the 58 non-*HLA* lead variants using 2,861 endpoints in FinnGen R6 PheWeb. In addition, we performed a lookup for SNPs within 1.5 Mb range of each lead SNP, noting previously reported lead SNPs in the GWAS catalogue. GWAS catalogue lead SNPs associated with blood markers and lab values were excluded from this annotation. We reported similar FinnGen disease endpoints together, so that e.g. “Childhood

asthma" and "Asthma as main diagnosis" are reported simply as "Asthma". For URD phenotypes, association with URD is not included.

Bayesian analysis of shared variant effects

In order to estimate the shared and distinct phenotypic impacts of specific loci in our GWAS results, we used a Bayesian framework, where we assessed for a shared effect between phenotypes. This framework adjusted for overlapping controls using a previously reported variance-based adjustment[67]. The framework considered three types of impact: none (the "null model"), unique ("one phenotype only"), and shared. Shared impact combined a hypothesized "fixed" model, where the variant has same effect size for all phenotypes, and a "correlated" model, where the variant has similar, but not necessarily the same, effects for all phenotypes. The posterior probability of the "fixed" and "correlated" models were added together and called the "shared" model when comparing with the null model and the unique effects models. The prior probability of "fixed" and "correlated" models were half of that of "null" and "one phenotype only" models so that each of the compared models (null, shared and each unique models) had the same prior probability. We interpreted a model with at least 75 % posterior probability as "most probable" model.

Shared impact between IURDs was analyzed in four tiers, corresponding to the four phenotype groups (IURDs, and sinonasal, pharyngeal, and allergic sinonasal diseases). The first tier analyzed heterogeneity of GWS SNPs identified in IURD GWAS based on their impact to CRNP, sinonasal diseases, pharyngeal diseases, and CLT. The second tier analyzed heterogeneity of GWS SNPs identified in the sinonasal disease GWAS on their impact to NSD

and CISD. The third tier analyzed the heterogeneity of GWS SNPs of the tonsil disease GWAS on CDTA and PA. The fourth and final tier analyzed the heterogeneity of GWS SNPs of the CISD GWAS on VAR, CRS and NP.

LD Score Regression

We estimated both the SNP-based observed-scale heritability and genetic correlation (r_g) by performing LD Score Regression using the LDSC toolset[47]. This method works by using an "LD score" to estimate the amount of linkage disequilibrium (LD) each SNP has with the rest of the genome under the polygenic model, and regresses the χ^2-statistic from a GWAS on the LD score, which also allows the estimation confounding bias. For our analysis, we used a previously calculated LD structure distributed by the ldsc.py software package. The distributed LD structure is based on the 1KG Phase 3 European dataset, and was merged to LD scores with the HapMap v3 variants[85]. The GWAS summary statistics were merged with 1,217,311 SNPs for which the LD scores was precalculated. 1,073 SNPs were removed due to being strand-ambiguous, 1,328 SNPS were removed due to duplicated rs-numbers, and 594 SNPs due to differing FinnGen and HapMap annotation. The remaining 1,190,282 SNPs were used in all FinnGen genetic correlation calculations.

LD score regression is developed for use in logistic and linear regression GWAS, while a GWAS using the SAIGE mixed model is not applicable for heritability estimates[45]. Therefore, the observed scale SNP-based heritability estimates were calculated using summary statistics from separate GWASs, in turn run using independent subsets for all phenotypes and using standard logistic regression. For the heritability analyses, a total of 54,784 SNPs were removed from the

initial 1,217,311 SNPs used for reference, and the observed scaled heritability estimate was calculated from the remaining 1,162,527 SNPs. In the logistic regression GWASs, we again used age, sex, PCs 1–10 and genotyping cohort as covariates.

We analyzed genetic correlation using LD Score regression (see below) to recognize shared heritability among IURD phenotypes. We grouped together phenotypes based on previously used thresholds, starting at $r_g > 75$ %. This grouped six of the eight IURD phenotypes into two groups: one group formed by VAR, CRS, NP, and NSD; another group being formed by CDTA and PA. Raising the threshold even further, to $r_g > 75$ %, distinguished a third group consisting of VAR, CRS, and NP. These groups were labelled "sinonasal diseases", "pharyngeal diseases" and "chronic inflammatory sinonasal diseases", respectively. We additionally analyzed genetic correlation to phenotypes detected in the phenome-wide analysis. Summary statistics for endpoints were from FinnGen release 6, with the exception of inflammatory bowel disease where a previously published analysis[86] was used.

Multi-marker Analysis of GenoMic Annotation (MAGMA)

We investigated gene- and gene set enrichment separately using Multi-marker Analysis of GenoMic Annotation (MAGMA[66]). Briefly, the pipeline calculates the mean χ^2 statistic from IURD GWAS summary statistics per gene, and thus obtains a p-value for the gene. MAGMA was analyzed using the FUMA pipeline that tests association for 19,535 curated genes; thus, the adjusted p-value threshold was set to $p < 0.05 / 19{,}535 = 2.56e\text{-}6$. Genes with p-value below this threshold were considered to associate with the relevant phenotype. We again employed the UK Biobank release 2 reference panel, with 1,000 randomly selected individuals for reference to

reduce runtime. Gene analysis was performed with default FUMA parameters, only considering SNPs that overlap genes. The *HLA* region was not omitted from MAGMA runs. We next tested for enrichment of genes involved in 4,761 curated gene sets and 5,917 GO terms included in the FUMA pipeline. Here the level of statistical significance was set with Bonferroni correction at $p < 0.05/(4,761+5,917) = 4.68e-6$.

Comparison with UK Biobank

We used the Pan-UKB dataset[51], based on the UK Biobank (UKB) participants, for replication of observed associations. FinnGen endpoints were matched to Pan-UKB phenotypes by finding the most similar Pan-UKB endpoint to each FinnGen endpoint. The similarity was calculated using the ICD-10 phenotype definitions as proxies. The FinnGen endpoints were mapped to corresponding ICD-10 endpoints. The UKB endpoints that were ICD-10 endpoints were used as is, and the UKB endpoints that were Phecode endpoints were mapped to ICD-10 endpoints using a Phecode-to-ICD-10 mapping. For each FinnGen and UKB phenotype, the similarity between them was calculated by counting the union and intersection of their sets of ICD-10 endpoints, and dividing the size of their ICD-10 intersection with the size of their ICD-10 union. A FinnGen and UKB pairing where this similarity score reached 1.0 was considered a perfect match. UKB variants were aligned to variants in FinnGen. Only phenotypes with a perfect match were included. In case of no exact match between SNPs (ref and alt differ between studies), matching was tried by flipping strand and/or switching ref->alt and alt->ref for the UKB variant. If there were multiple variants in the same position, the exact match was favored.

Comparison was performed between FinnGen R6 GWAS summary statistics and lifted (hg19->hg38) UKB GWAS summary statistics (EUR population). Non-codirectional effects (product of betas less than 0) were considered non-replicated. For the remaining summary statistics, a test statistic was calculated with the formula

$$z = \frac{(\beta_1 - \beta_2)^2}{SE_1^2 + SE_2^2}$$

where β_i is the effect size of study i, and SE_i is the standard error of the effect estimate in study i. The test statistic z was assumed to follow a χ^2 distribution with one degree of freedom.

TABLES

Abbr.	Phenotype	ICD-10	Cases	λ_{GC}	Loci	CS
IURD	Inflammatory and infectious upper respiratory diseases	J30-9	61,197	1.2038	20	21
VAR	Vasomotor and allergic rhinitis	J30	8,975	1.0772	6	5
CRNP	Chronic rhinitis, nasopharyngitis and pharyngitis	J31	6,518	1.0354	0	0
CRS	Chronic rhinosinusitis	J32	10,435	1.0864	7	5
NP	Nasal polyps	J33	3,919	1.0618	13	12
NSD	Nasal septal deviation	J34.2	7,716	1.0584	1	1
CDTA	Chronic diseases of tonsils and adenoids	J35	29,135	1.1908	21	22
PA	Peritonsillar abscess	J36	4,863	1.0527	5	5
CLT	Chronic laryngitis and laryngotracheitis	J37	2,623	1.0204	0	0
DPPT	Diseases of pulp and periapical tissues	K04	48,687	1.1175	2	2
CP	Chronic periodontitis	K05.30-1	14,631	1.0610	1	1

Table 1: Description of genome-wide association studies. The cases were identified using registry data from hospitals and specialized out-patient clinics. Abbr. = Abbreviation for phenotype. The same set of controls (n = 199,208) was used in all IURD GWAS (ICD-10 category J3). Control counts were 211,718 for DPPT and 245,774 for CP. λ_{GC} is the genomic inflation factor. 'Loci' is the number of genome-wide significant loci (incl HLA). CS is the number of credible sets from fine-mapping with at least one GWS SNP (the HLA region was not fine-mapped).

					J30-J39: Inflammatory and infectious upper respiratory diseases (n = 61,197)					
							Sinonasal dis.	J31: CRNP	Pharyngeal dis.	J37: CL
BAND	**Alleles**	**Nearest gene**	**rsid**	**EAF**	**OR (95 % CI)**	***p***	**OR (95 % CI)**	**OR (95 % CI)**	**OR (95 % CI)**	**OR (95 %**
2q12.1	G/T	*IL18R1,IL1RL1*	rs950881	16.2 %	0.95 (0.92–0.97)	2.69E-08	0.90 (0.87–0.93)	0.96 (0.91–1.02)	0.97 (0.94–1.00)	1.02 (0.94–
2q13	A/G	*MIR4435-2HG*	rs1045267	66.0 %	0.95 (0.93–0.97)	6.26E-11	0.99 (0.96–1.01)	0.99 (0.95–1.04)	0.91 (0.88–0.93)	0.96 (0.90–
2q33.3	A/G	*ADAM23*	rs189411872	1.3 %	1.20 (1.12–1.28)	1.87E-08	1.13 (1.03–1.24)	1.02 (0.86–1.21)	1.40 (1.28–1.53)	1.00 (0.77–
3q13.12	G/T	*CCDC54*	rs187688516	1.7 %	0.85 (0.80–0.91)	3.72E-08	0.89 (0.82–0.97)	0.89 (0.77–1.03)	0.83 (0.76–0.90)	0.82 (0.66–
3q21.2	C/T	*SLC12A8*	rs1980080	65.3 %	0.96 (0.94–0.98)	1.52E-08	0.98 (0.95–1.00)	0.96 (0.92–1.00)	0.93 (0.91–0.96)	1.01 (0.95–
4q24	A/G	*NFKB1*	rs4648051	32.0 %	0.95 (0.93–0.97)	4.88E-10	0.96 (0.93–0.98)	0.98 (0.94–1.03)	0.94 (0.91–0.96)	1.00 (0.93–
4q24	G/GC	*TET2*	rs5860793	71.6 %	1.05 (1.03–1.07)	1.24E-09	1.03 (1.01–1.06)	1.06 (1.02–1.11)	1.08 (1.05–1.11)	1.06 (0.99–
5p13.2	T/TT	*IL7R*	rs11406102	33.0 %	0.96 (0.94–0.98)	2.12E-08	0.95 (0.92–0.97)	0.97 (0.93–1.01)	0.95 (0.93–0.98)	0.95 (0.89–
5q22.1	T/C	*TSLP*	rs1837253	75.4 %	1.05 (1.03–1.07)	6.66E-09	1.10 (1.07–1.13)	1.04 (0.99–1.09)	1.03 (1.01–1.06)	1.03 (0.96–
6p22-21	-/-	*HLA*	-	-	1.07 (1.04–1.09)	3.87E-12	1.11 (1.08–1.14)	0.57 (0.41–0.78)	1.16 (1.10–1.21)	1.13 (1.06–
9p24.1	T/C	*IL33*	rs2095044	76.0 %	0.95 (0.93–0.97)	3.34E-10	0.88 (0.85–0.91)	0.96 (0.91–1.00)	0.99 (0.97–1.02)	0.97 (0.90–
9q33.3	C/T	*NEK6*	rs3758213	38.1 %	0.95 (0.93–0.97)	1.75E-10	0.94 (0.92–0.97)	0.94 (0.90–0.98)	0.95 (0.93–0.98)	0.99 (0.93–
12p13.31	A/G	*LTBR*	rs10849448	75.5 %	0.94 (0.92–0.96)	7.68E-15	0.97 (0.94–1.00)	1.00 (0.96–1.05)	0.90 (0.88–0.92)	1.03 (0.96–
12p13.31	A/G	*PZP*	rs967884772	1.1 %	1.23 (1.14–1.32)	1.15E-08	1.24 (1.12–1.37)	1.00 (0.83–1.20)	1.22 (1.11–1.35)	1.27 (0.95–
13q21.33	T/C	*KLHL1*	rs9542155	64.3 %	0.96 (0.94–0.98)	4.60E-08	0.98 (0.95–1.01)	1.00 (0.96–1.04)	0.94 (0.92–0.96)	0.96 (0.90–
14q21.1	A/G	*FBXO33*	rs112672184	0.2 %	1.57 (1.34–1.83)	8.73E-09	1.66 (1.32–2.07)	1.22 (0.81–1.83)	1.72 (1.40–2.12)	1.16 (0.62–
15q22.33	C/T	*SMAD3*	rs17293632	26.2 %	1.05 (1.02–1.07)	2.63E-08	1.07 (1.04–1.10)	1.04 (0.99–1.08)	1.04 (1.01–1.06)	1.07 (1.00–
17p11.2	C/G	*TNFRSB13B*	rs573841223	2.5 %	1.21 (1.15–1.27)	1.41E-15	1.04 (0.97–1.12)	1.12 (0.99–1.26)	1.42 (1.33–1.52)	0.98 (0.81–
17q21.2	A/G	*KRT19*	rs551772399	2.1 %	1.15 (1.09–1.22)	2.08E-08	1.16 (1.08–1.25)	1.11 (0.98–1.27)	1.17 (1.10–1.26)	0.91 (0.73–
19p13.3	G/A	*ZBTB7A*	rs74178437	73.8 %	1.05 (1.03–1.07)	2.29E-08	1.03 (1.01–1.06)	1.04 (1.00–1.09)	1.07 (1.04–1.10)	1.06 (0.99–

Table 2: Lead SNPs of Genome-wide associated loci in GWAS of IURD (n = 61,197). Phenotypes (sinonasal diseases, CRNP, pharyngeal diseases and CL) have effect sizes annotated with 95 % CI. 'Alleles' denotes the reference / effect alleles. EAF: Effect allele frequency. The *HLA* locus is annotated for the lead SNP according to each phenotype, not the corresponding SNP of the parent GWAS (IURD).

LOCUS	GWAS	RSID	OR	95 % CI	*p*	EAF	FE	Gene	CONSEQ
2q12.1	CRS	rs1041973	0.90	(0.87–0.94)	6.21E-08	20.5 %	0.91	*IL1RL1*	missense
4q24a	IURD	rs2272676	0.96	(0.94–0.98)	7.84E-09	34.9 %	1.10	*NFKB1*	splice donor
4q24b	CDTA	rs2454206	0.94	(0.92–0.96)	5.59E-09	34.1 %	0.89	*TET2*	missense
5p13.2	IURD	rs6897932	0.96	(0.94–0.98)	3.26E-08	33.1 %	1.28	*IL7R*	missense
5q31.1b	NP	rs1050152	0.86	(0.81–0.91)	3.73E-09	31.7 %	0.72	*SLC22A4*	missense
16p12.1	CISD	rs144651842	0.89	(0.85–0.94)	7.68E-08	7.8 %	**119**	*IL4R*	missense
16q22.1	CDTA	rs117556162	0.88	(0.83–0.92)	2.92E-08	4.7 %	0.86	*CARMIL2*	missense
17p11.2	CDTA	rs72553883	1.43	(1.33–1.53)	2.39E-26	2.4 %	**3.52**	*TNFRSF13B*	missense
17q11.2	PHAR. DIS.	rs2071587	1.09	(1.06–1.14)	3.05E-08	9.2 %	1.02	*FOXN1*	missense
17q12	CRS	rs3744374	0.90	(0.86–0.93)	9.94E-10	23.0 %	0.95	*GAS2L2*	missense
17q21.1	CISD	rs11557467	0.94	(0.91–0.96)	1.60E-08	55.9 %	1.14	*ZPBP2*	missense
17q21.1	CISD	rs2305479	0.94	(0.91–0.96)	2.69E-08	54.8 %	1.14	*GSDMB*	missense
17q21.2	IURD	rs201961848	1.15	(1.08–1.22)	5.25E-06	1.4 %	***INF***	*G6PC*	missense
19q13.33	CRS	rs602662	1.08	(1.05–1.12)	1.45E-07	41.4 %	0.79	*FUT2*	missense
19q13.43	CDTA	rs774674736	2.01	(1.58–2.55)	9.62E-09	0.018 %	***INF***	*ZNF417*	frameshift
21q22.13	PA	rs2230033	1.11	(1.06–1.16)	1.62E-06	48.9 %	0.90	*KCNJ15*	missense

Table 3: 16 functional variants in protein-coding genes included in 95 % credible sets with at least one GWS SNP. GWAS column denotes the genome-wide association study where the variant is identified. Odds ratios and p-values are with regard to the phenotype in GWAS column; the most specific phenotype is represented if the variant appears in several GWAS. EAF = Effect allele frequency. FE = Enrichment in FinnGen (Finnish-enriched variants are bolded), i.e. allele frequency compared with non-Finnish participants in gnomAD. CONSEQ = most severe consequence annotated with the Variant Effect Predictor (VEP). PHAR.DIS. = Pharyngeal diseases. INF = no non-Finnish carriers in gnomAD.

Phenotype	Gene set	Genes	β	p_{adj}
CRS	MHC class II receptor activity	9	1.6939	0.036904
NP	MHC class II receptor activity	9	2.5209	1.96E-07
NP	MHC class II receptor complex	14	1.7043	0.000367
NP	Positive regulation of interleukin 13 production	15	1.2187	0.000453
NP	Regulation of interleukin 4 production	28	0.90161	0.003912
NP	Interleukin 13 production	24	0.88367	0.004649
NP	MHC protein complex	22	1.0476	0.008563
NP	Interleukin 4 production	34	0.78389	0.011623
CDTA	Cytokine mediated signaling pathway	748	0.16085	0.005677
CDTA	TNFR2 pathway	16	1.0288	0.00569
CDTA	Reactome cytokine signaling in immune system	835	0.14811	0.006434
CDTA	Peptidyl serine autophosphorylation	8	1.3374	0.027797
PA	Negative regulation of morphogenesis of an epithelium	16	1.1272	0.025657

Table 4: Gene sets enriched in MAGMA analyses. Gene sets were identified based on genes identified as phenotype-associated in MAGMA analysis. β is the effect size in the MAGMA gene set enrichment analysis. The considered gene sets encompassed a set 4761 curated gene sets and 5917 Gene Ontology terms, as used in the FUMA pipeline. The *p*-value has been adjusted for these sets (10,678 tests).

FIGURES

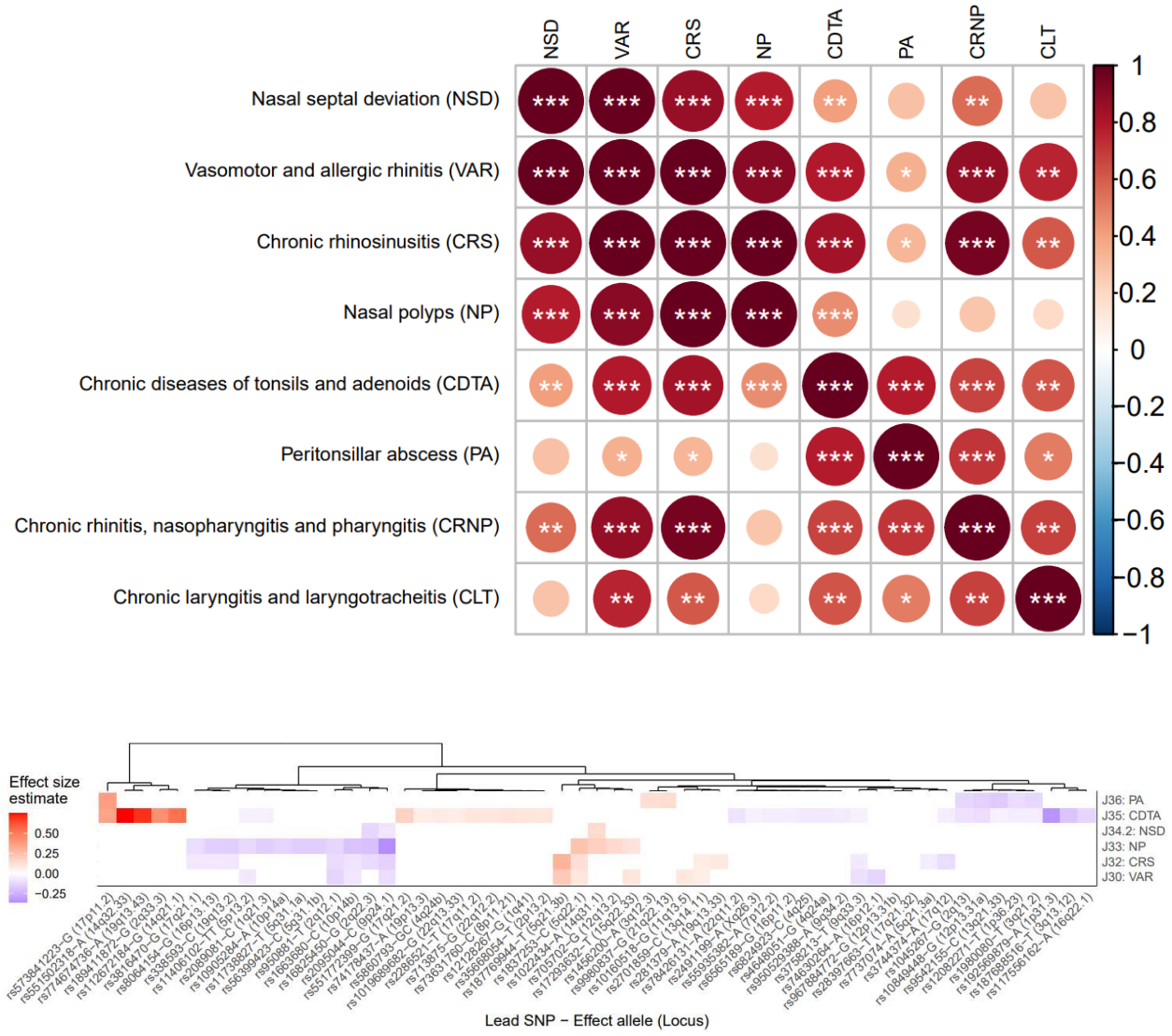

Figure 1: Genome-wide association and genetic correlation structure of inflammatory and infectious upper respiratory diseases (IURDs). A (top): Genetic correlation of IURDs distinguishing vasomotor and allergic rhinitis (VAR), chronic rhinosinusitis (CRS), and nasal polyposis (NP) as a near-completely genetically correlated cluster. The color of the circle indicates genetic correlation with red indicating positive correlation and blue indicating negative correlation. **B (bottom):** Effect sizes of 57 genetic loci across six IURD phenotypes. Not included are the *HLA* locus and 2p16.1, which strongly associated only with NSD (β = 2.27). Red indicates a positive effect size estimate, and blue indicates a negative effect size estimate. Only significant associations (p < 0.01) are shown. SNPs on the y-axis are ordered according to a clustering algorithm (denoted above and detailed in Methods). The clusters show shared genetic heritability within the recognized phenotype groups of sinonasal (NSD, VAR, CRS, NP) and pharyngeal diseases (CDTA, PA). The locus 2p16.1 had a strong NSD-specific effect that was beyond the scale plotted here. There were no associated loci for CRNP or CL. *p<0.05, **p<0.005, ***p<8.4e-5

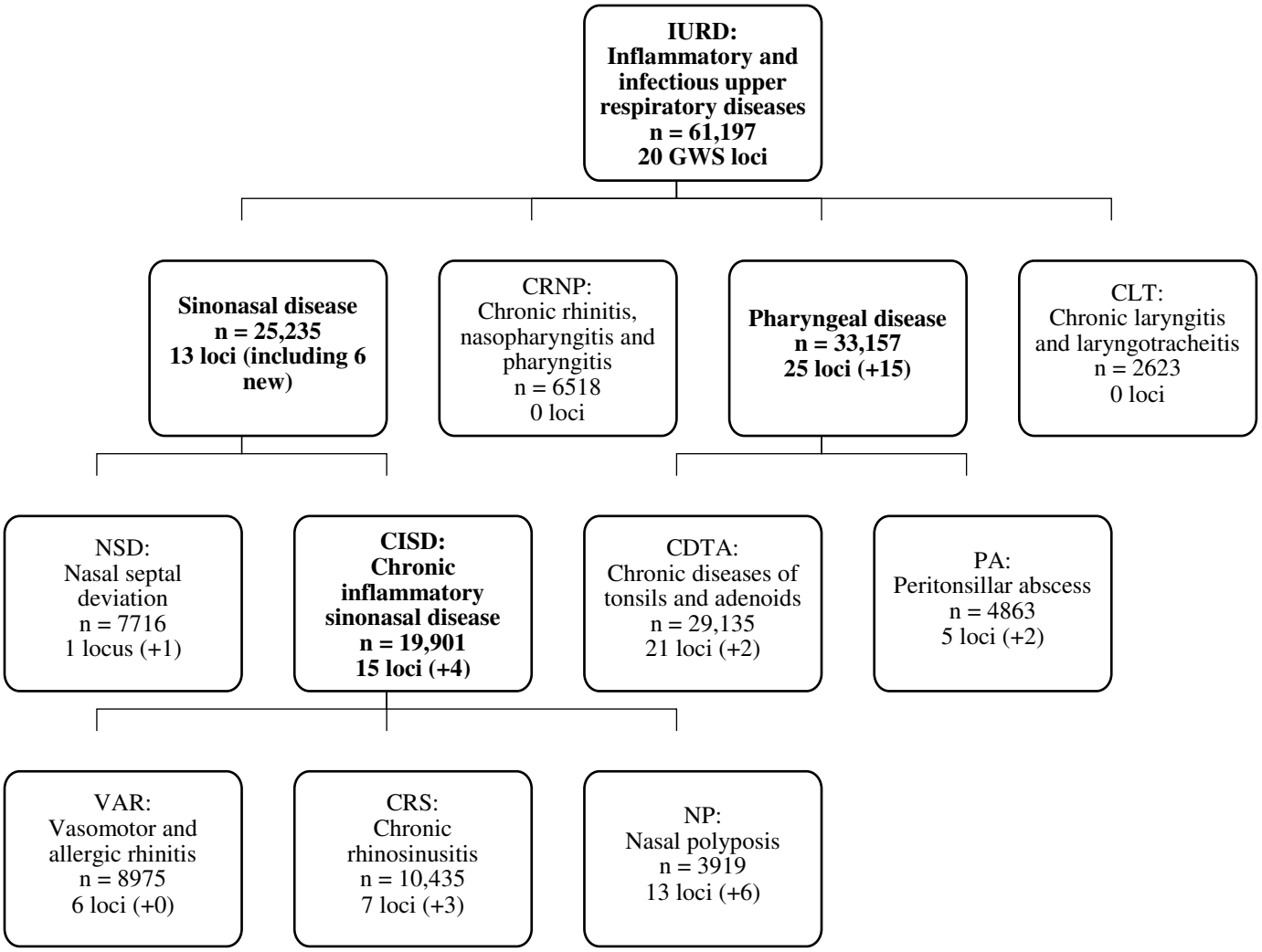

Figure 2: The IURD phenotype structure, based on genetic correlation between phenotypes. The boxes represent a GWAS of a IURD phenotype or group, stating the name, case count and number of GWS loci, indicating in parenthesis the GWS loci that were not detected in any directly antecedent ("parent") GWAS. E.g., there were three loci in CRS that had not been detected in CISD, sinonasal disease or IURD GWAS. The hierarchical structure shows the phenotypes included in the parent phenotype. The IURD GWAS also included ICD codes J38 and J39 (not depicted); for these, no separate GWAS was performed. Sinonasal disease phenotypes (NSD, VAR, CRS and NP) had a genetic correlation 78 % or higher, as estimated using LD Score regression. Pharyngeal diseases (CDTA and PA) had a genetic correlation of 79 %. The observed genetic correlations between chronic inflammatory sinonasal diseases (VAR, CRS, and NP) were 90 % or higher.

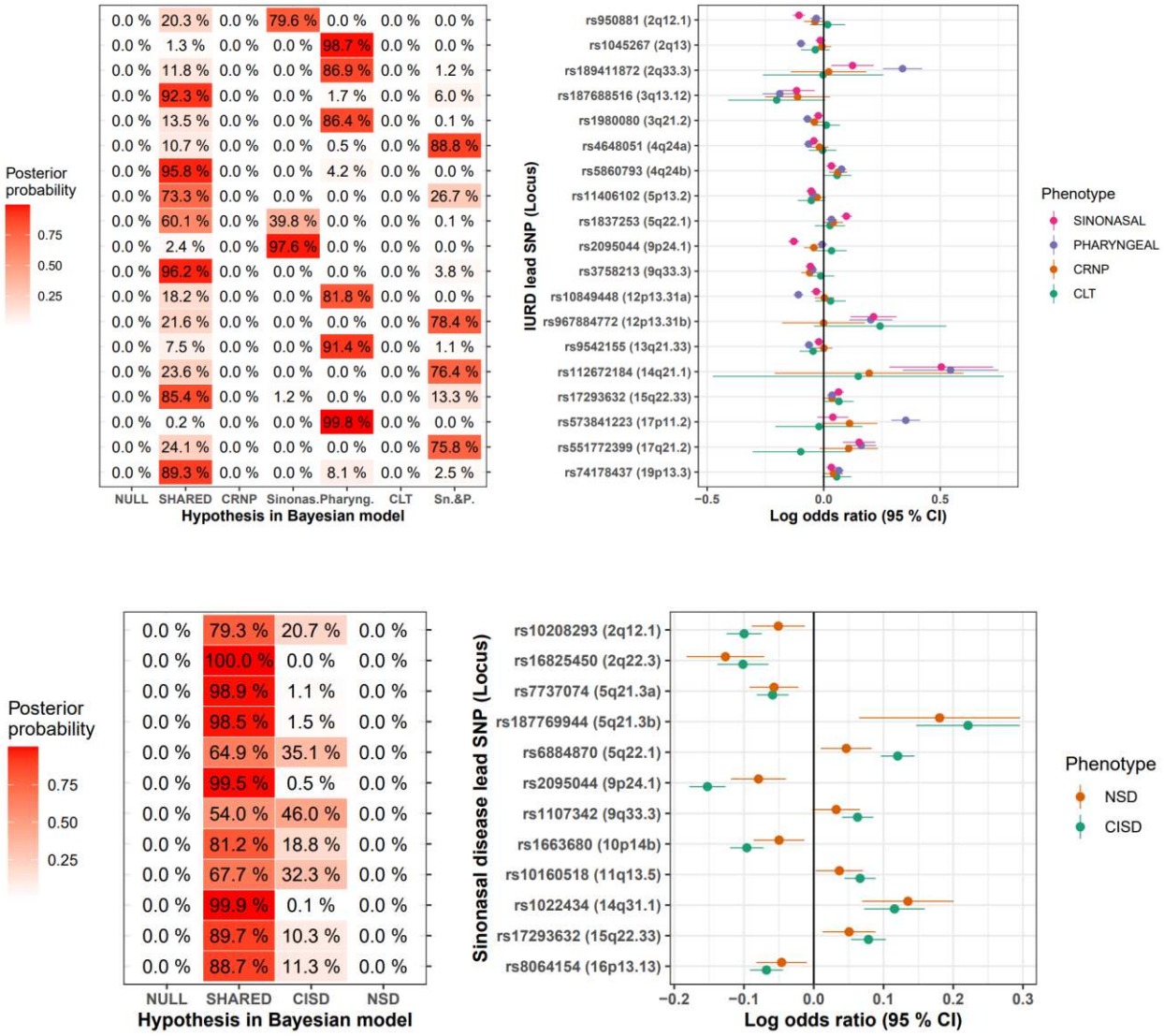

Figure 3: Shared impact between phenotypes for IURD lead SNPS and sinonasal disease lead SNPs. **A (upper left)**: Bayesian posterior probabilities of hypothetical models displayed on the x-axis for IURD lead SNPs (y-axis; locus in parenthesis). Models correspond to NULL = null model; SHARED = Fixed or correlated effect model across phenotypes (CRNP, sinonasal disease,pharyngeal disease and CL); CRNP = CRNP only; Sinonas. = sinonasal disease only; Pharyng. = pharyngeal disease only; CLT = CLT only; Sn.&P. = Sinonasal AND pharyngeal disease (fixed effect). **B (upper right)**: Odds ratios of IURD lead SNPs for phenotypes, with 95 % confidence intervals. **C (lower left)**: Bayesian posterior probabilities of hypothetical models for each lead SNP from the sinonasal disease GWAS. Models as in panel A; additionally CISD = CISD diseases only; and NSD = NSD only. **D (lower right)**: Odds ratios of sinonasal disease lead SNPs for the two phenotypes.

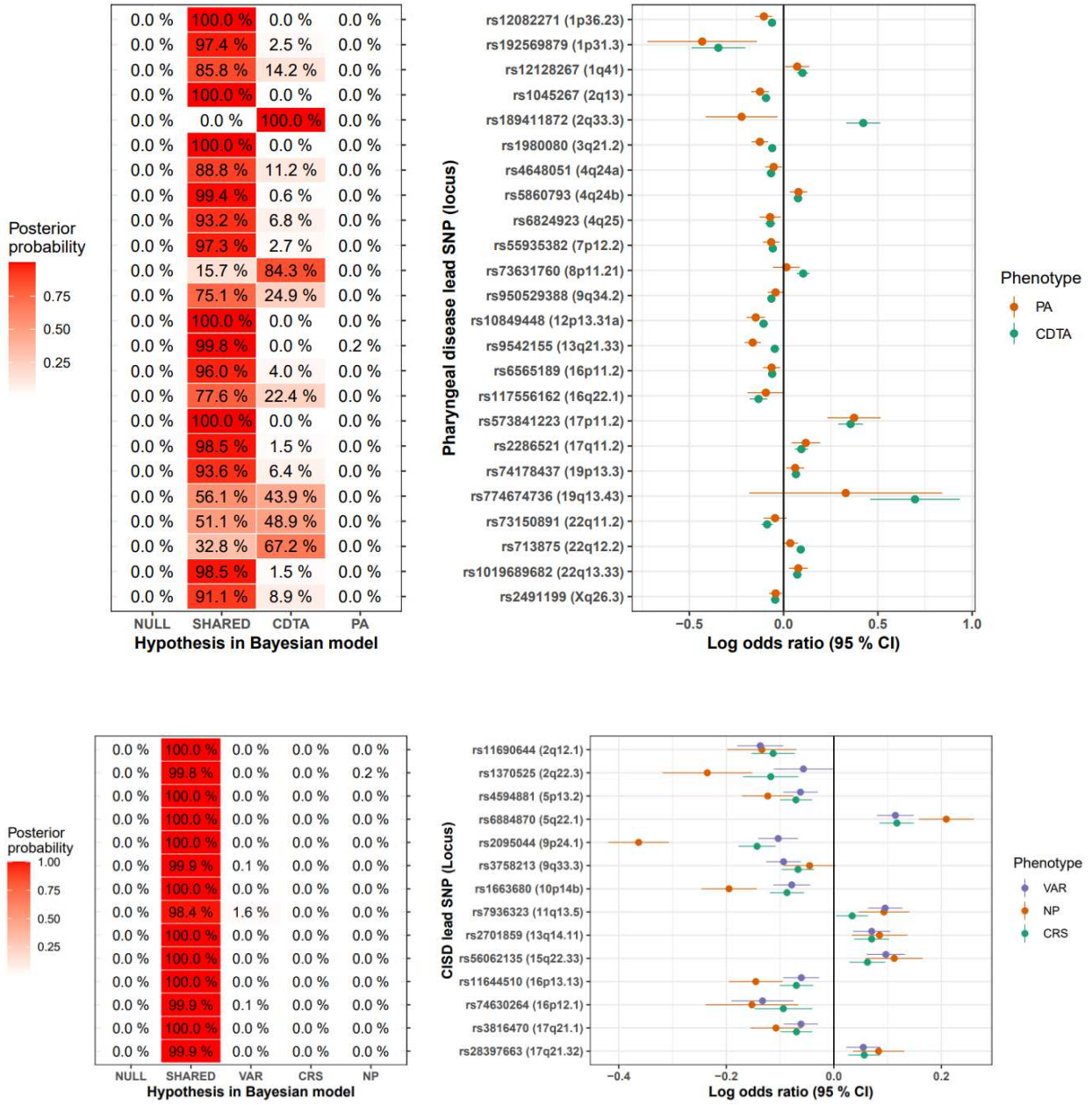

Figure 4: Shared impact between phenotypes for tonsil and sinus disease lead SNPS. **A (upper left)**: Bayesian posterior probabilities of hypothetical models displayed on the x-axis for pharyngeal disease lead SNP (y-axis). Models correspond to NULL = null model; CORR = Fixed or correlated effect model across phenotypes (CDTA and PA); CDTA = CDTA only; PA = PA only. **B (upper right)**: Odds ratios of pharyngeal lead SNPs for CDTA and PA, with 95 % confidence intervals. **C (lower left)**: Bayesian posterior probabilities of hypothetical models for each lead SNP from the sinus disease GWAS. Models NULL and SHARED as in panel A; additionally VAR = VAR only; CRS = CRS only; NP = NP only. **D (lower right)**: Odds ratios of sinus disease lead SNPs for the three phenotypes VAR, CRS, and NP.

Supplementary Materials

REFERENCES

1 Bachert, C., Bhattacharyya, N., Desrosiers, M. & Khan, A. H. Burden of Disease in Chronic Rhinosinusitis with Nasal Polyps. *J Asthma Allergy* **14**, 127-134, doi:10.2147/JAA.S290424 (2021).

2 Kværner, K. J., Nafstad, P. & Jaakkola, J. J. K. Upper Respiratory Morbidity in Preschool Children: A Cross-sectional Study. *Archives of Otolaryngology–Head & Neck Surgery* **126**, 1201-1206, doi:10.1001/archotol.126.10.1201 (2000).

3 Sánchez Choez, X. *et al.* Medical Cost of Upper Respiratory Tract Infections in Children in Ambulatory Care. *Value in Health Regional Issues* **26**, 1-9, doi:10.1016/j.vhri.2020.10.001 (2021).

4 Khaltaev, N. & Axelrod, S. J. J. o. T. D. Chronic respiratory diseases global mortality trends, treatment guidelines, life style modifications, and air pollution: preliminary analysis. *2019* **11**, 2643-2655 (2019).

5 Staikūnienė, J., Vaitkus, S., Japertienė, L. M. & Ryškienė, S. Association of chronic rhinosinusitis with nasal polyps and asthma: Clinical and radiological features, allergy and inflammation markers. **44**, 257 (2008).

6 Fokkens, W. J. *et al.* European Position Paper on Rhinosinusitis and Nasal Polyps 2020. *Rhinology* **58**, 1-464, doi:10.4193/Rhin20.600 (2020).

7 Kim, S. Y., Kim, H.-R., Min, C. & Choi, H. G. Bidirectional association between asthma and otitis media in children. *Allergy, Asthma & Clinical Immunology* **17**, 7, doi:10.1186/s13223-020-00500-7 (2021).

8 Ji, J., Sundquist, J. & Sundquist, K. Tonsillectomy associated with an increased risk of autoimmune diseases: A national cohort study. *Journal of Autoimmunity* **72**, 1-7, doi:10.1016/j.jaut.2016.06.007 (2016).

9 Shih, L.-C. *et al.* Chronic rhinosinusitis and premorbid autoimmune diseases: a population-based case–control study. *Scientific Reports* **10**, 18635, doi:10.1038/s41598-020-75815-x (2020).

10 Janszky, I., Mukamal, K. J., Dalman, C., Hammar, N. & Ahnve, S. Childhood appendectomy, tonsillectomy, and risk for premature acute myocardial infarction—a nationwide population-based cohort study. *European Heart Journal* **32**, 2290-2296, doi:10.1093/eurheartj/ehr137 (2011).

11 Maccioni, L. *et al.* Obesity and risk of respiratory tract infections: results of an infection-diary based cohort study. *BMC Public Health* **18**, 271, doi:10.1186/s12889-018-5172-8 (2018).

12 Chapman, S. J. & Hill, A. V. S. Human genetic susceptibility to infectious disease. *Nature Reviews Genetics* **13**, 175-188, doi:10.1038/nrg3114 (2012).

13 Burgner, D., Jamieson, S. E. & Blackwell, J. M. Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *The Lancet Infectious Diseases* **6**, 653-663, doi:10.1016/S1473-3099(06)70601-6 (2006).

14 Yazdani, R. *et al.* Common variable immunodeficiency: epidemiology, pathogenesis, clinical manifestations, diagnosis, classification, and management. **30**, 14-34 (2020).

15 Tangye, S. G. *et al.* Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. *Journal of Clinical Immunology* **40**, 24-64, doi:10.1007/s10875-019-00737-x (2020).
16 Beck, D. B. & Aksentijevich, I. Susceptibility to severe COVID-19. *Science* **370**, 404-405, doi:10.1126/science.abe7591 (2020).
17 Ovsyannikova, I. G., Haralambieva, I. H., Crooke, S. N., Poland, G. A. & Kennedy, R. B. The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. **296**, 205-219, doi:10.1111/imr.12897 (2020).
18 The COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European Journal of Human Genetics* **28**, 715-718, doi:10.1038/s41431-020-0636-6 (2020).
19 Andersén, H. *et al.* Dyspnea has an association with lifestyle: differences between Swedish and Finnish speaking persons in Western Finland. *European Clinical Respiratory Journal* **8**, 1855702, doi:10.1080/20018525.2020.1855702 (2021).
20 Christensen, S. H. *et al.* A clear urban–rural gradient of allergic rhinitis in a population-based study in Northern Europe. *European Clinical Respiratory Journal* **3**, 33463, doi:10.3402/ecrj.v3.33463 (2016).
21 Qiu, H. *et al.* The burden of overall and cause-specific respiratory morbidity due to ambient air pollution in Sichuan Basin, China: A multi-city time-series analysis. *Environmental Research* **167**, 428-436, doi:10.1016/j.envres.2018.08.011 (2018).
22 Joshi, M., Goraya, H., Joshi, A. & Bartter, T. Climate change and respiratory diseases: a 2020 perspective. **26**, 119-127, doi:10.1097/mcp.0000000000000656 (2020).
23 Olin, A. *et al.* Stereotypic Immune System Development in Newborn Children. *Cell* **174**, 1277-1292.e1214, doi:10.1016/j.cell.2018.06.045 (2018).
24 Xu, L., Earl, J., Bajorski, P., Gonzalez, E. & Pichichero, M. E. Nasopharyngeal microbiome analyses in otitis-prone and otitis-free children. *International Journal of Pediatric Otorhinolaryngology* **143**, 110629, doi:10.1016/j.ijporl.2021.110629 (2021).
25 Hong, S.-N. *et al.* Chronic rhinosinusitis with nasal polyps is associated with chronic otitis media in the elderly. *European Archives of Oto-Rhino-Laryngology* **274**, 1463-1470, doi:10.1007/s00405-016-4363-0 (2017).
26 Cingi, C. *et al.* Multi-morbidities of allergic rhinitis in adults: European Academy of Allergy and Clinical Immunology Task Force Report. *Clinical and Translational Allergy* **7**, 17, doi:10.1186/s13601-017-0153-z (2017).
27 Tiotiu, A. *et al.* Current opinions for the management of asthma associated with ear, nose and throat comorbidities. *European Respiratory Review* **27**, 180056, doi:10.1183/16000617.0056-2018 (2018).
28 Kim, J. H. *et al.* Association between the sinus microbiota with eosinophilic inflammation and prognosis in chronic rhinosinusitis with nasal polyps. *Experimental & Molecular Medicine* **52**, 978-987, doi:10.1038/s12276-020-0458-1 (2020).
29 Barnes, K. C. Evidence for common genetic elements in allergic disease. *Journal of Allergy and Clinical Immunology* **106**, S192-S200, doi:10.1067/mai.2000.110150 (2000).

30 Braunstahl, G.-J. The unified immune system: Respiratory tract–nasobronchial interaction mechanisms in allergic airway disease. *Journal of Allergy and Clinical Immunology* **115**, 142-148, doi:10.1016/j.jaci.2004.10.041 (2005).
31 von Mutius, E. & Smits, H. H. Primary prevention of asthma: from risk and protective factors to targeted strategies for prevention. *The Lancet* **396**, 854-866, doi:10.1016/S0140-6736(20)31861-4 (2020).
32 Gliklich, R. E. & Metson, R. The health impact of chronic sinusitis in patients seeking otolaryngologic care. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* **113**, 104-109, doi:10.1016/S0194-59989570152-4 (1995).
33 Dykewicz, M. S. *et al.* Rhinitis 2020: A practice parameter update. *Journal of Allergy and Clinical Immunology* **146**, 721-767, doi:10.1016/j.jaci.2020.07.007 (2020).
34 Bousquet, J. *et al.* Allergic rhinitis. *Nature Reviews Disease Primers* **6**, 95, doi:10.1038/s41572-020-00227-0 (2020).
35 Sugita, K. & Kabashima, K. Tight junctions in the development of asthma, chronic rhinosinusitis, atopic dermatitis, eosinophilic esophagitis, and inflammatory bowel diseases. **107**, 749-762, doi:10.1002/JLB.5MR0120-230R (2020).
36 Hsu, J. *et al.* Genetics of chronic rhinosinusitis: State of the field and directions forward. *Journal of Allergy and Clinical Immunology* **131**, 977-993.e975, doi:10.1016/j.jaci.2013.01.028 (2013).
37 Dan, J. M. *et al.* Recurrent group A *Streptococcus* tonsillitis is an immunosusceptibility disease involving antibody deficiency and aberrant T_{FH} cells. *Science Translational Medicine* **11**, eaau3776, doi:10.1126/scitranslmed.aau3776 (2019).
38 Haapasalo, K. *et al.* The Psoriasis Risk Allele *HLA-C*06:02* Shows Evidence of Association with Chronic or Recurrent Streptococcal Tonsillitis. *Infection and Immunity* **86**, e00304-00318, doi:10.1128/IAI.00304-18 (2018).
39 Kristjansson, R. P. *et al.* A loss-of-function variant in ALOX15 protects against nasal polyps and chronic rhinosinusitis. *Nature Genetics* **51**, 267-276, doi:10.1038/s41588-018-0314-6 (2019).
40 Tian, C. *et al.* Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nature Communications* **8**, 599, doi:10.1038/s41467-017-00257-5 (2017).
41 Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics* **48**, 709-717, doi:10.1038/ng.3570 (2016).
42 Ferreira, M. A. *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nature Genetics* **49**, 1752-1757, doi:10.1038/ng.3985 (2017).
43 Lees, J. A. *et al.* Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. *Nature Communications* **10**, 2176, doi:10.1038/s41467-019-09976-3 (2019).
44 DeLorenze, G. N. *et al.* Polymorphisms in HLA Class II Genes Are Associated With Susceptibility to Staphylococcus aureus Infection in a White Population. *The Journal of infectious diseases* **213**, 816-823, doi:10.1093/infdis/jiv483 (2015).
45 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341, doi:10.1038/s41588-018-0184-y (2018).

46 Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. **82**, 1273-1300, doi:10.1111/rssb.12388 (2020).
47 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).
48 Hung, S.-H., Tsai, M.-C., Lin, H.-C. & Chung, S.-D. Allergic Rhinitis Is Associated With Periodontitis: A Population-Based Study. **87**, 749-755, doi:10.1902/jop.2016.150539 (2016).
49 Byun, S. H. *et al.* Increased Risk of Chronic Periodontitis in Chronic Rhinosinusitis Patients: A Longitudinal Follow-Up Study Using a National Health-Screening Cohort. *Journal of Clinical Medicine* **9**, 1170, doi:10.3390/jcm9041170 (2020).
50 Craig, J. R. *et al.* Diagnosing odontogenic sinusitis: An international multidisciplinary consensus statement. *International Forum of Allergy & Rhinology* **n/a**, 1-14, doi:10.1002/alr.22777 (2021).
51 Pan-UKB team. *Pan-ancestry genetic analysis of the UK Biobank*, <https://pan.ukbb.broadinstitute.org> (2020).
52 Savenije, O. E. M. *et al.* Interleukin-1 receptor-like 1 polymorphisms are associated with serum IL1RL1-a, eosinophils, and asthma in childhood. *Journal of Allergy and Clinical Immunology* **127**, 750-756.e755, doi:10.1016/j.jaci.2010.12.014 (2011).
53 Qiu, R., Zhao, H., Wang, A., Gong, Y. & Liu, Q. Association of genetic variants in chromosome 17q21 and adult-onset asthma in a Chinese Han population. *BMC Medical Genetics* **12**, 133, doi:10.1186/1471-2350-12-133 (2011).
54 Parulekar, A. D., Kao, C. C., Diamant, Z. & Hanania, N. A. Targeting the interleukin-4 and interleukin-13 pathways in severe asthma: current knowledge and future needs. *Current Opinion in Pulmonary Medicine* **24**, 50-55, doi:10.1097/mcp.0000000000000436 (2018).
55 Castro, M. *et al.* Dupilumab Efficacy and Safety in Moderate-to-Severe Uncontrolled Asthma. *The New England journal of medicine* **378**, 2486-2496, doi:10.1056/NEJMoa1804092 (2018).
56 Xuan, C. *et al.* Association between OCTN1/2 gene polymorphisms (1672C-T, 207G-C) and susceptibility of Crohn's disease: a meta-analysis. *International Journal of Colorectal Disease* **27**, 11-19, doi:10.1007/s00384-011-1265-x (2012).
57 McGovern, D. P. B. *et al.* Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Human molecular genetics* **19**, 3468-3476, doi:10.1093/hmg/ddq248 (2010).
58 Salzer, U. *et al.* Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet* **37**, 820-828, doi:10.1038/ng1600 (2005).
59 Castigli, E. *et al.* TACI is mutant in common variable immunodeficiency and IgA deficiency. *Nat Genet* **37**, 829-834, doi:10.1038/ng1601 (2005).
60 Pulvirenti, F. *et al.* Clinical Associations of Biallelic and Monoallelic *TNFRSF13B* Variants in Italian Primary Antibody Deficiency Syndromes. *Journal of Immunology Research* **2016**, 8390356, doi:10.1155/2016/8390356 (2016).

61 Fliegauf, M. *et al.* Haploinsufficiency of the NF-κB1 Subunit p50 in Common Variable Immunodeficiency. *The American Journal of Human Genetics* **97**, 389-403, doi:10.1016/j.ajhg.2015.07.008 (2015).
62 Puel, A., Ziegler, S. F., Buckley, R. H. & Leonard, W. J. Defective IL7R expression in T-B+NK+ severe combined immunodeficiency. *Nature Genetics* **20**, 394-397, doi:10.1038/3877 (1998).
63 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).
64 Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330, doi:10.1126/science.aaz1776 (2020).
65 Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715.e1716, doi:10.1016/j.cell.2018.10.022 (2018).
66 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).
67 Bhattacharjee, S. *et al.* A Subset-Based Approach Improves Power and Interpretation for the Combined Analysis of Genetic Association Studies of Heterogeneous Traits. *The American Journal of Human Genetics* **90**, 821-835, doi:10.1016/j.ajhg.2012.03.015 (2012).
68 Vivier, E. *et al.* Innate Lymphoid Cells: 10 Years On. *Cell* **174**, 1054-1066, doi:10.1016/j.cell.2018.07.017 (2018).
69 Lambrecht, B. N., Hammad, H. & Fahy, J. V. The Cytokines of Asthma. *Immunity* **50**, 975-991, doi:10.1016/j.immuni.2019.03.018 (2019).
70 Ordovas-Montanes, J. *et al.* Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649-654, doi:10.1038/s41586-018-0449-8 (2018).
71 Zhu, Z. *et al.* A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nature Genetics* **50**, 857-864, doi:10.1038/s41588-018-0121-0 (2018).
72 Torgerson, D. G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nature Genetics* **43**, 887-892, doi:10.1038/ng.888 (2011).
73 Han, Y. *et al.* Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nature Communications* **11**, 1776, doi:10.1038/s41467-020-15649-3 (2020).
74 Verstraete, K. *et al.* Structure and antagonism of the receptor complex mediated by human TSLP in allergy and asthma. *Nature Communications* **8**, 14937, doi:10.1038/ncomms14937 (2017).
75 Dijk, F. N. *et al.* Genetic regulation of *IL1RL1* methylation and IL1RL1-a protein levels in asthma. *European Respiratory Journal* **51**, 1701377, doi:10.1183/13993003.01377-2017 (2018).
76 Das, S. *et al.* GSDMB induces an asthma phenotype characterized by increased airway responsiveness and remodeling without lung inflammation. *Proceedings of the National Academy of Sciences* **113**, 13132-13137, doi:10.1073/pnas.1610433113 (2016).

77 Saikumar Jayalatha, A. K., Hesse, L., Ketelaar, M. E., Koppelman, G. H. & Nawijn, M. C. The central role of IL-33/IL-1RL1 pathway in asthma: From pathogenesis to intervention. *Pharmacology & Therapeutics* **225**, 107847, doi:10.1016/j.pharmthera.2021.107847 (2021).
78 de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature Genetics* **49**, 256-261, doi:10.1038/ng.3760 (2017).
79 Hayden, M. S. & Ghosh, S. NF-κB in immunobiology. *Cell Research* **21**, 223-244, doi:10.1038/cr.2011.13 (2011).
80 Ahmad, S. *et al.* The Key Role of TNF-TNFR2 Interactions in the Modulation of Allergic Inflammation: A Review. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.02572 (2018).
81 The Severe Covid-19 GWAS Group. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *The New England journal of medicine* **383**, 1522-1534, doi:10.1056/NEJMoa2020283 (2020).
82 Zhang, H., Mooney, C. J. & Reilly, M. P. ABO Blood Groups and Cardiovascular Diseases. *International Journal of Vascular Medicine* **2012**, 641917, doi:10.1155/2012/641917 (2012).
83 Ruotsalainen, S. E. *et al.* An expanded analysis framework for multivariate GWAS connects inflammatory biomarkers to functional variants and disease. *European Journal of Human Genetics* **29**, 309-324, doi:10.1038/s41431-020-00730-8 (2021).
84 Institute for Molecular Medicine Finland (FIMM). *Sequencing Initiative Suomi project (SISu) v3*, <http://sisuproject.fi> (2020).
85 Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861, doi:10.1038/nature06258 (2007).
86 Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* **47**, 979-986, doi:10.1038/ng.3359 (2015).

Acknowledgments: The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme Corp, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, and Novartis AG. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta) and Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi). Finnish Biobank Cooperative -FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (https://site.fingenious.fi/en/) managed by FINBB. We thank the GWAS catalog (https://www.ebi.ac.uk/gwas/), the GTEx database (https://www.gtexportal.org/), the DICE database (https://dice-database.org/), the Pan-UK database (https://pan.ukbb.broadinstitute.org/), and Tian and colleagues, for sharing expression data and summary statistics of GWAS studies.

Funding: This study was funded by the Sigrid Juselius Foundation, and the Horizon 2020 Research and Innovation Programme [grant number 667301 (COSYN) to A.P.], the National Institute of Health (Grant no. 1U01MH105666-01), the Finnish Medical Foundation (Grant no. 3264), the Centre of Excellence Complex Disease Genetics (CoECDG, University of Helsinki, Academy of Finland Grant no. 312074 for A.P., Grant no. 312062 for S.R., Grant no. 312073 for J.K. and Grant no. 312075 for M.D.), and the Doctoral School for Population Health (University of Helsinki). A.S.H was supported by the Academy of Finland, Grant no. 321356 and S.R. was also supported by the Finnish Foundation for Cardiovascular Research and University of Helsinki HiLIFE Fellow and Grand Challenge grants. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions: Each author's contribution(s) to the paper is listed according to the CRediT model.

Conceptualization: ECS, JTR, TK, HH, MP, SRi, MD, TP, AM, AP

Methodology: MK, MP, HH, JM, JK

Validation: JM

Formal analysis: ECS, JK, JM, SRu

Investigation: JK, FINNGEN

Resources: FINNGEN

Data Curation: JK, TK, ASH, FINNGEN

Visualization: ECS

Funding acquisition: AP, AM, TP, MD, SRi, MK, MP, ASH

Project administration: AP

Supervision: AP, AM, TP, MD, SRi, MK, MP, ASH

Writing – original draft: ECS

Writing – review & editing: JK, JTR, ASH, HH, MP, SRi, MD, TP, AM, AP, STS

Competing interests. STS reports consultancies for AstraZeneca, ERT, Novartis, Sanofi Pharma, and Roche Products and a grant of GSK, outside the submitted work. All other authors declare that they have no competing financial interests.

Data and materials availability: FINNGEN summary-level data (release 6) will be made publicly available in Aug 2021 (https://www.finngen.fi/). Software used in this analysis are publicly available software distributed by the respective websites (SAIGE: https://www.leelabsg.org/software; SuSiE: https://github.com/stephenslab/susieR; LDSC: https://github.com/bulik/ldsc/) with developer pages on github. The Bayesian analysis framework is publicly detailed in the cited work, in addition to the website (https://github.com/trochet/metabf).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementalmaterialsv1.2.docx
- Supplementarytablesv1.2.xlsx
- Table2IURDloci.xlsx