De novo identification of complex multimorbid conditions by integration of gene regulation and protein interaction networks with genome-wide association studies

Roan Zaied
Liggins Institute, The University of Auckland https://orcid.org/0000-0001-5981-9940
Tayaza Fadason
University of Auckland https://orcid.org/0000-0002-1170-7883
Justin O'Sullivan (justin.osullivan@auckland.ac.nz)
Liggins Institute, The University of Auckland https://orcid.org/0000-0003-2927-450X

Article

Keywords: asthma, SNP function, multimorbidity, prediction, network analysis

Posted Date: March 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1313207/v1

Abstract

Network analysis represents a powerful approach for untangling the relationships between complex diseases. Here, we integrated information on physical contacts between common single nucleotide polymorphisms (SNPs; MAF≥0.05) and expressed genes (TPM>0.1), with expression quantitative trait loci (eQTL) data from whole blood to construct a blood-specific spatial gene regulatory network (GRN). Using this GRN, we located the genes that are functionally affected by asthma-associated SNPs to identify the asthma-specific module. We then expanded outwards to identify curated protein-protein interactions occurring 1 to 4 edges away from the asthma module. The eQTLs spatially regulating the identified genes were used to interrogate the GWAS Catalog to identify enriched traits (hypergeometric test; FDR≤0.05) within each of the 4 levels. This led to *de novo* identification of traits both previously reported (e.g., sarcoidosis and lung cancer) and unreported (e.g., disc-degeneration) to be multimorbid with asthma, and pinpoints the variants and genes bridging asthma to the multimorbidities. Our discovery approach identifies significantly enriched traits in the regulatory space proximal to a disease of interest, in the tissue of interest, without *a priori* selection of the interacting diseases. The predictions it makes expand our understanding of possible shared therapeutic targets, especially for diseases such as asthma where no cure is currently available.

Introduction

Asthma is a complex inflammatory airway disease characterized by reversible air-flow obstruction and airway remodeling (1–3). In 2019, an estimated 262 million cases of asthma were reported, and it was associated with around 416 thousand deaths (4). Asthma is a heterogenous phenotype with a wide range of interactions and presentations with other phenotypes (5). Multimorbidity with complex diseases (e.g. eczema, mood disorders, and diabetes) has been previously reported suggesting the existence of shared molecular pathways (6–9). Consistent with this, the genetics of asthma are complex and the functional impacts of known asthma-associated variants are poorly understood (10). A large meta-analysis of genome-wide association studies (GWAS) conducted by the Trans-National Asthma Genetic Consortium (TAGC) identified 878 SNPs in 18 loci as being associated with asthma (11). These loci were enriched for enhancer marks, particularly in immune cell types, suggesting roles in immune regulation (11). The specific causal mechanisms of disease-associated SNPs (da-SNPs) are, however, challenging to identify and their association with the observed multimorbidity is yet to be elucidated (12). Identifying the regulatory mechanisms that link asthma to its multimorbid conditions will have a significant impact on our understanding of disease-disease interactions and pave the way for potential novel therapeutic approaches.

Assigning functions to da-SNPs is a difficult undertaking because most da-SNPs fall in non-coding regions and have small individual effect sizes (12). Moreover, expression quantitative trait loci (eQTL) analyses have demonstrated that long-range regulatory interactions occur at a significant number of loci, making the conventional nearest-gene-approach for identification of da-SNP functional targets insufficient (13–15). We have previously demonstrated the use of chromatin interaction (Hi-C) data

coupled with eQTL analysis for the identification of target genes whose regulation requires physical interaction with the eQTL-tagged locus (16, 17). These spatially constrained gene regulatory networks represent one way of assigning functions to da-SNPs by linking variants to the genes they regulate.

Network-based analysis provides an opportunity to investigate the inter-relationships of complex polygenic diseases within network neighborhoods. As such, network-based analyses represent a possible step-change in our ability to understand the multimorbidity, development, and therapeutic avenues for pathologically related diseases (18–20). In network studies, the understanding that biologically similar diseases segregate in modules within the same network neighborhood has been used in attempts to identify conditions that are associated with asthma (20, 21). In such studies, *a priori* selection of the interacting conditions enabled the identification of: 1) overlapping genes between asthma and nine immune-mediated diseases (20); and 2) coding genes that link diseases with recognized similarities (e.g. chronic obstructive pulmonary disease (COPD) and asthma) (22). Unfortunately, these approaches require *a priori* assumptions that limit investigations to conditions whose pathophysiological relationships are clinically recognized.

In this study, we constructed a spatial gene regulatory network (GRN) comprised of genes that are expressed in whole blood and whose transcript levels are associated with spatially constrained eQTLs involving all common SNPs (MAF≥0.05). Using this GRN network we located the genes that are functionally affected by asthma-associated SNPs. Translation of the blood regulatory network into a protein-protein interaction network (PPIN) enabled the identification of proteins that interact with the core asthma gene network. We subsequently identified sub-networks that are separated by 1-4 edges away from the asthma disease module and identified the complex traits that are enriched within those modules. This agnostic approach enabled us to *de novo* identify diseases that are associated with asthma such as systemic lupus erythematosus (SLE), sarcoidosis, and lung cancer. These disease have been previously observed to be multimorbid with asthma (23–26). We also identified traits that are proximal to the asthma disease module whose association with asthma are yet to be confirmed (e.g., disc degeneration and hip circumference). Collectively, our results demonstrate the utility of spatial GRNs in the *de novo* identification of complex traits that are associated with a disease of interest.

Materials And Methods

Data sources

- Expression levels of genes expressed in whole blood, measured in median transcripts per million (TPM; Supplementary Data 1), were obtained from GTEx (https://gtexportal.org/home/datasets/GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz, 15/3/2021).
- All variants genotyped from whole blood samples were downloaded from GTEx (https://gtexportal.org/home/datasets/GTEx_Analysis_2017-06-05_v8_WholeGenomeSeq_838Indiv_Analysis_Freeze.lookup_table.txt, 4/01/2021).

- Single nucleotide polymorphisms (SNPs) and their associated traits were downloaded from the GWAS Catalog (https://www.ebi.ac.uk/gwas/docs/file-downloads/, 18/06/2021).
- Gene biotype information was obtained from GENCODE (http://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_26/gencode.v26.annotation.gtf.gz, 05/06/2021).
- gnomAD was used to retrieve gene-level constraint metrices (Supplementary Data 2; https://gnomad.broadinstitute.org/downloads#v2-constraint/gnomad.v2.1.1.lof_metrics.by_gene.txt, 11/01/2021).
- Gene start and end positions were obtained from GTEx (https://gtexportal.org/home/datasets/gencode.v26.GRCh38.genes.gtf, 17/02/2021).
- Cell type specific eQTLs were obtained from DICE (https://dice-database.org/downloads/, 09/02/2021).
- GRCh38.p13 is the human genome build used in this study.

Identification of the whole blood spatial gene regulatory network (GRN)

The CoDeS3D pipeline (https://github.com/Genome3d/codes3d-v2) (27) was used to identify spatial eQTLs. In brief, Hi-C data was downloaded and Hi-C libraries constructed for four primary blood cell lines (Supplementary Data 3), as described in Rao et al. (28). The same restriction enzymes used to prepare the libraries (HindIII and MboI) were used to digitally digest the GRCh38 genome into DNA fragments. SNPs (MAF $\geq$0.05) present within whole blood samples (GTEx (15)) were used to interrogate the Hi-C libraries for interactions captured between a DNA fragment overlapping a gene (as defined by gencode v.26) and the DNA fragment containing the queried SNP (Fig. 1). Only SNP-gene interactions captured in more than one Hi-C cell line replicate were included in downstream analysis. SNP-gene pairs were then tested to identify eQTLs within whole blood (GTEx (15)). The GTEx eQTL calculator was used to identify spatial trans-eQTLs. Finally, multiple testing correction (Benjamini-Hochberg) was performed and spatial eQTLs with an adjusted p-value $\leq$0.05 were selected as being significant. A SNP that: 1) physically contacts; and 2) is at least partially associated with the transcription level of the interacting gene, is termed a spatial eQTL, while the gene is termed a spatial eGene. Spatial cis-eQTLs regulate genes that are <1 Mb away from the eQTL on the same chromosome, while trans-intrachromosomal eQTLs regulate genes that are >1 Mb away on the same chromosome. Trans-interchromosomal spatial eQTLs regulate genes that reside on different chromosomes.

Construction of the blood protein-protein interaction network (PPIN)

eGenes that were expressed in whole blood (TPM >0.1 and $\geq$6 reads in a minimum of 20% of tested samples) were identified by querying the GTEx Catalog (accessed 15/3/2021). These eGenes were used to interrogate the STRINGdb package (29) (v. 2.2.2) for potential protein-protein interactions having a medium confidence score of $\geq$ 0.700 (Supplementary Data 4). The sources used as evidence of interaction include text-mining, experiments, databases, co-expression, and genomics context prediction

channels *i.e.*, fusion, neighborhood, and gene co-occurrence. Conversions between Ensemble gene identifiers and Ensemble protein identifiers were performed using the biomaRt (30) package.

Construction of disease-specific modules

383 asthma-associated SNPs were identified using a keyword search for the exact term “Asthma” in the GWAS Catalog (v1.0.2, 18/06/2021; Supplementary Data 5). The asthma-associated SNP set was then used to subset the asthma GRN (Supplementary Data 6) and PPIN (Supplementary Data 7) from the larger blood GRN and PPIN, respectively. The resulting sub-network represented the asthma (“+0” level) PPIN. Cytoscape was used to visualize the asthma PPIN (31). The same process was repeated for acute lymphoblastic leukemia (ALL) using ALL-associated SNPs obtained using a keyword search for “Acute lymphoblastic leukemia (childhood)” (Supplementary Data 8).

Identification of traits that are proximal to the asthma and ALL disease modules using the blood PPIN

The disease-specific modules (asthma and ALL) were expanded to include curated protein interactions (STIRNG) that fall between 1 and 4 edges from the core disease network. For example, first-neighbor blood PPIN proteins share 1 edge (*i.e.,* directly interact) with the disease-specific “+0” PPIN and form the PPIN on the “+1” level. Similarly, second-neighbor proteins share 1 edge with proteins on the “+1” level (and a maximum of 2 edges with the core disease PPIN) and form the “+2” PPIN (excluding proteins found on the “+0” and the “+1” level). eQTLs that regulate the expression of the protein encoding genes on each of the levels were subsequently identified by querying the blood GRN. eQTLs for the protein encoding genes on the +1, +2, +3, or +4 levels were queried against the GWAS Catalog (v1.0.2, 18/06/2021) to identify traits that had been associated with these variants through GWAS. The hypergeometric distribution test was used to identify enrichment of traits at each level. Multiple testing correction was done and traits with an adjusted Benjamini-Hochberg (32) p-values ≤ 0.05 were considered to be significantly enriched.

Gene Ontology analysis

Gene Ontology (GO) enrichment was performed using the g:GOSt module of the R package g:Profiler (33). The eGenes within the asthma GRN were tested for enrichment within reactome pathways (34), biological process, cellular compartment, and molecular function GO terms. Multiple testing correction was performed (Benjamini-Hochberg) (32) and adjusted p-values ≤ 0.05 were deemed significant.

Results

The whole blood spatial GRN is enriched for genes intolerant to loss of function (LOF)

SNPs (MAF$\geq$ 0.05) within the GTEx dataset (n=$40x10^6$) were downloaded (dbGaP accession: phs000424.v8.p2; approved project number: #22937) and screened for spatial eQTLs in whole blood

(Fig. 1) using the CoDeS3D algorithm (27). The resulting whole blood GRN is comprised of 1,713,885 spatial eQTLs (1,077,379 SNPs and 14,871 eGenes expressed in whole blood). Of those, 1,634,655 eQTLs are cis-acting (eQTL-eGene are separated by <1Mb), 67,014 are trans-intrachromosomal (eQTL-eGene pair are separated by ≥1Mb on the same chromosome), and 12,216 are trans-interchromosomal spatial eQTLs (eQTL-eGene pair are located on different chromosomes; Supplementary Figs. 1a and 1b). eGenes regulated in cis have significantly more ($p < 0.0001$) eQTL interactions than those regulated through trans-interchromosomal and trans intrachromosomal interactions (mean ± SD, 126 ± 2.4, 9.8 ± 1.3 and 4.1 ± 0.9, respectively) and eGenes that are subject to trans-intrachromosomal regulation have significantly more ($p < 0.0001$) eQTLs than eGenes regulated by trans-interchromosomal eQTLs (Fig. 2a). eGenes whose transcript levels are associated with cis-acting eQTLs also have significantly ($p < 0.0001$) higher expression levels than those regulated in trans (Fig. 2b). Notably, we observed that eGenes associated with trans-acting eQTLs tend to be enriched for genes that are intolerant to loss of function (LOF) mutations, as estimated by the LOEUF score (35) (Fig. 2c).

Gene regulation patterns in mammalian systems have a strong developmental component. Therefore, we hypothesized that cis and trans-regulatory elements would be involved in the fine-tuning and regulation of transcript levels for most genes expressed in blood. Notably, of the genes expressed in whole blood (median TPM >0.1), 13,674 were associated with spatial eQTLs, while 5,415 genes were not (Supplementary Fig. 1c). There was no significant difference between the expression levels of protein-coding eGenes and non-eGenes. However, eGenes were significantly more intolerant to LOF than protein coding non-eGenes (Supplementary Figs. 1d and 1e). Correlating the number of spatial eQTLs against the number of GTEx SNPs per chromosome shows chromosomes 4, 8, 9, 13 and 18 to have fewer spatial eQTLs than expected and chromosomes 6, 17 and 19 to have more spatial eQTLs than expected, with chromosome 6 having the greatest deviation from the prediction (Fig. 2d).

The human leucocyte antigen (HLA) region in chromosome 6 has a spike of spatial eQTL-eGene interactions

We observed that chromosome 6 was enriched for spatial eQTL-eGene interactions – given the number of variants that are mapped to the chromosome. There was a notable enrichment of cis interactions in gene-dense regions across Chromosome 6, when compared to gene-poor regions (Fig. 3a). We observed that there was a significant increase in cis-interactions about the major histocompatibility locus (chr6: 29690551, 33102442) on chromosome 6, this is the most gene-dense region across chromosome 6 (Fig. 3b), and it has a well-recognized role in immune responses. The HLA locus aligns with a near 2-fold increase in the ratio of eQTL-eGene interactions to SNPs in whole blood (Fig. 3c). This spike in whole blood eQTLs across the HLA locus was not observed in other tissues (*i.e.*, left ventricular and adult brain cortical tissues; Supplementary Fig. 2). Tissue-specific spikes in ratios of eQTL-eGene interactions to SNPs were also observed within other chromosomes in other tissues (Supplementary Fig. 2a-c), including chromosomes that were depleted for eQTLs within whole blood (Supplementary Fig. 2b).

Related immune cell types share similar cell specific gene regulatory networks

Whole blood contains subsets of different nucleated cell types. Therefore, we obtained cell type-specific cis-eQTL data to determine the relative contributions of the different cell classes to the whole blood GRN. The Database of Immune Cell Expression, Expression of quantitative trait loci and Epigenomics (DICE) (see methods; Data Sources) contained 1,189,347 unique cis-acting eQTLs from across 15 immune cell types. These include innate immune cell types (natural killer cells [NK], classical monocytes, and non-classical monocytes [M2 cells]), naïve adaptive immune cells (naïve B, naïve $CD4^+$ T cells, activated naïve CD4+ T cells [CD4_stim], naïve $CD8^+$ T cells, activated naïve CD8+ T cells [CD8_stim], and naïve T_{REG} cells), and $CD4^+$ T memory subtypes (Memory $T_{REG,}$ T_H1, T_H 1/17 [T_H star], T_H17, T_H2,, and follicular helper T cells [T_{FH}]) (36). 19.41% of the 1,713,885 spatial eQTLs within our whole blood GRN were also identified within DICE (DICE spatial eQTLs) (Fig. 4a). Of those, 14,418 spatial eQTLs were shared across all 15 cell types. The normalized effect sizes of the shared spatial eQTL-eGene pairs were used to identify the correlating immune cell types (Fig. 4b). Moreover, unsupervised hierarchal clustering of the normalized effect size of the 14,418 shared spatial eQTLs identifies distinct cell type clusters sharing similar spatial gene regulatory pathways in whole blood (Fig. 4c).

Immune cell types sharing similar asthma GRN suggests their shared regulatory roles in the etiology of asthma

We used asthma-associated SNPs to identify the asthma GRN within the whole blood GRN. The asthma GRN consists of 302 eQTL-eGene interactions between 122 eQTLs and 132 eGenes (Supplementary Data 6). Gene ontology enrichment analysis of the asthma-associated eQTL-targeted eGenes identified enrichment within immune-related biological processes (Supplementary Fig. 3a). Similarly, analysis of molecular functions and reactome pathways identified enrichment in immune-related processes and in lipid metabolism (e.g., linoleoyl-CoA desaturase activity, carnitine transmembrane transporter activity, and linoleic acid (LA) metabolism etc.; Supplementary Figs. 3b and 3c). Lastly, we hypothesized that cell-specific asthma GRNs (curated by integrating eQTL data from DICE) could identify inter-cellular associations that are informative of asthma development. We thus performed unsupervised hierarchal clustering of the z-score-transformed normalized effect size of all asthma-associated spatial eQTL-eGene interactions in the 15 immune cells from DICE. This identified immune cell clusters sharing related asthma spatial GRNs such as T_H17, T_H2, and Memory T_{REG} cells suggesting their potential similar regulatory roles in the etiology of asthma (Fig. 5a).

The blood GRN identified conditions that are multimorbid with asthma

We hypothesized that the asthma-associated eQTL-targeted genes interacted with each other as part of a biological process. Therefore, we translated the asthma spatial GRN into an asthma PPIN that consisted of 89 proteins (nodes) having 63 connections (edges) (Fig. 5b and Supplementary Data 7). The asthma PPIN included HLA proteins and other proteins previously reported to be associated with asthma e.g., PDCD1 (82, 83), PRKCQ (37), GSDMA (38) and GSDMB (39).

Various complex traits such as airway hyperresponsiveness, metabolic syndrome, and impaired lung function have been identified as multimorbid conditions in asthma patients suggesting the existence of shared molecular pathways contributing to their development (40). Therefore, we identified the proteins that are interacting with the asthma PPIN and used them to determine the traits that are proximal to asthma in regulatory space. This was achieved by identifying the proteins within the blood PPIN that are connected to the asthma PPIN by less than 4 edges. The 1,963 interacting proteins that were in the first interacting level (+1 edge from the core asthma PPIN) within the extended asthma PPIN were regulated by 186,436 eQTLs within the blood GRN. In the +2 level, 525,136 eQTLs targeted transcript levels for 5,534 eGenes. In the +3 level, 246,626 eQTLs targeted transcript levels for 2,113 eGenes. Finally, in the "+4" level, 31911 eQTLs targeted transcript levels for 256 eGenes. Querying the GWAS Catalog (v1.0.2, 18/06/2021) identified 13 enriched (hypergeometric test; FDR≤0.05) traits in the "+1" level, 1 trait in the "+2" level, 5 traits in the "+3" level and 4 traits in the "+4" level (Fig. 6; Supplementary Data 9 and 10).

The same analysis was repeated using the ALL PPIN which consisted of 8 proteins having 4 edges. Expanding four levels away from the ALL PPIN identified 250 proteins on the +1 level, 3,536 +2 proteins, 4,764 +3 proteins and 1,275 +4 proteins. Enrichment analysis (hypergeometric test; FDR≤0.05) of the eQTLs regulating the expression of the genes on each of the 4 levels identified 39 enriched GWAS traits in the +1 level, 13 traits in the +2 level, 5 traits in the +3 level and 8 traits in the +4 level (Supplementary Fig. 4).

Discussion

Untangling the relationships between complex phenotypes can prove elusive because the molecular and biological mechanisms that influence their presentation interact across multiple levels of biological information. In this study, we performed a *de novo* discovery analysis of conditions that are multimorbid with asthma using a whole blood spatial GRN. In effect, the whole blood GRN acted as a Rosetta stone connecting the asthma-associated SNPs with eQTLs from an extended set of curated protein-protein interactions. Thus, we were able to use molecular interactions to identify genetic variants that affect traits known to be multimorbid with asthma as well as other traits having an unconfirmed association with asthma. Our *de novo* discovery approach can be applied to other complex polygenic disorders and will help to identify known and unknown interacting phenotypes, thereby providing information on the genetic variation and biological mechanisms that are potentially responsible for the interaction.

This study is not without limitations. Primary amongst them is that associations can only be identified for traits previously investigated in GWAS studies. An additional consequence of this limitation is that not all variations will be represented due to GWAS participants being primarily of European ancestry (41). Second, our method relies on regulatory connections having known protein interactions. eQTLs whose target genes do not form protein-protein interactions, or those forming unknown interactions, will thus be missed by our analysis. The third limitation is that the blood GRN is not dynamic, i.e., it represents a snapshot in time of the captured interactions and is thus subject to change. Fourth, our analysis combines information across multiple levels of biological organization that do not originate from the

same biological samples (i.e., eQTL data from GTEx, Hi-C datasets, population studies [GWAS], and protein-protein interaction). Notwithstanding these limitations, the ability to discover multimorbid traits *de novo* through described molecular pathways and by integrating resources across biological levels provides a potential step-change in our ability to identify why and how genetic variation contributes to complex phenotypes.

Asthma is a complex inflammatory airway disease whose features include airway remodeling and respiratory obstruction (1, 3). Consistent with these observations, we identified genetic variants that are associated with forced expiratory volume (FEV) and forced vital capacity (FVC) related phenotypes, which characterize respiratory functions, as being directly connected to protein encoding genes that are affected by asthma-associated SNPs. In addition, we also observed direct connections to proteins that are targeted by eQTLs associated with conditions such as Alzheimer's disease (AD), blood protein levels (i.e. interleukin-18 (IL-18) (42), macrophage inflammatory protein-1b (MIP-1b) (43, 44) and 90K protein (45, 46), etc.), SLE, and sarcoidosis, all of which are conditions known to be multimorbid with asthma. For instance, metanalysis studies report an increased risk for the development of SLE in asthmatics (pooled odds ratio (OR): 1.37; 95% CI 1.14–1.65; I2 = 67%) (23). Additionally, the genetic risk of sarcoidosis, which is a granulomatous multisystem disease of unknown etiology and pathology that predominantly affects the lungs (47, 48), had significant pleiotropy with the genetic risk of asthma (R^2=2.03%; p=8.89×10^{-9}) (26). These findings suggest a potential role played by the remaining identified traits in the multimorbidity of asthma.

The strength of the network approach we outlined is that it provides molecular insights for interactions between asthma multimorbidities. For example, despite reported links between asthma and an increased risk for AD (hazard ratio (HR): 2.62; 95% CI 1.71–4.02) (49), there are no known mechanisms through which this could occur. Our analysis, however, implicates 19 genes and 56 eQTLs in the asthma-AD multimorbid interaction. These include *ERCC1* (spatially regulated by the AD-associated eQTL, rs55923289), which interacts with both KAT5 and XRCC6 in the asthma PPIN. While no population studies associate *ERCC1* with AD, *ERCC1* mutant and knockout mice were reported to present with age-dependent neuronal pathology and cognitive decline (50, 51). Moreover, AD patients were reported to have decreased mRNA levels of *ERCC1* compared to healthy controls suggesting its role in AD pathology (52). Similarly for lung cancer, the risk of its development increases in adults with active asthma (HR 1.29; 95% CI 0.95–1.75) (24). This could be mediated through mechanisms such as chronic inflammation, which is associated with higher risk of developing cancer (53), structural and functional changes of the lungs caused by chronic asthma exacerbations e.g., thickening of the bronchial wall, fibrosis, and formation of scar tissue (54, 55), and asthma-associated molecular disruptions that contribute to the development of lung cancer (22). Consistent with this, our findings implicate 13 genes, which are spatially regulated by 24 eQTLs, as contributors to the observed multimorbidity. One of those is the amino peptidase *CTSH* produced by lung macrophages and is spatially regulated by the lung cancer-associated eQTLs rs4886591, rs11639372, and rs28408315. *CTSH* interacts with 8 different asthma eGenes (TNFSF4, HLA-DRB5, HLA-DQA2, HLA-DQB2, CHIT1, HLA-DRB1, HLA-DQA1 and HLA-DQB1) and

has been previously associated with lung cancer (61). Interestingly, *CTSH* was found to be overexpressed in both smoking lung cancer patients (63) and asthmatics (62). Collectively, our results identify the potential regulatory mechanisms contributing to the observed asthma multimorbidity and we propose that this should be subjected to further empirical investigation.

Conditions that have been identified in our analysis but were not reported to be associated with asthma, or their association with asthma is inconclusive, include disc degeneration, hip circumference (56, 57), mean arterial pressure, blood pressure related to alcohol consumption, and leprosy. Of those, we predict that traits found closer to the asthma PPIN (e.g., disc degeneration identified in the +1 level) to have stronger relationships with asthma. *TAP1* is one of the genes that we identified to be modulating the interaction between the asthma PPIN (through interactions with HLA-B) and disc degeneration-associated eQTLs (11 eQTLs). *TAP1* mediates the unidirectional translocation of peptide antigens across the endoplasmic reticulum for loading into MHC I molecules (58). Interestingly, previous studies report it to be associated with lumbar disc degeneration (59), asthma, and other inflammatory conditions such as rhinitis and dermatitis (60). Moreover, “TAP binding” was one of the enriched molecular functions of the asthma PPIN proteins (Supplementary Fig. 3b). These findings suggest potential increased risks for asthma-disc degeneration multimorbidities mediated through genes involved in inflammatory pathways such as *TAP1*. Asthma multimorbidities were also reported for traits found in the “+3” level including glycated hemoglobin levels (61), plasma-free amino acid levels (62), and linoleic acid (63, 64), as well as traits found in the “+4” level such as circulating chemerin levels (65). Thus, while proximity of the identified traits to the asthma module suggests stronger associations with asthma, higher level associations could also contribute to the risk of multimorbidity.

To further demonstrate the applicability of our approach to other complex disorders, we used it to identify the conditions that are proximal to ALL in regulatory space. Some of the conditions that we identified were previously reported to be associated with ALL (e.g., nephropathy, asthma (OR = 1.43, 95% CI: 1.10, 1.85) (66), disc degeneration (67), etc...) Drug associations were found to be common contributors to the observed multimorbidity. For instance, anticancer treatment in ALL has been associated with kidney function decline (68) and administration of immunosuppressive agents in systemic sclerosis patients has been associated with increased risk of hematological cancer (69). For disc degeneration, 90% of children with ALL are reported to have decreased bone mineralization, a contributor to disc degeneration (67, 70). Our analysis implicates 9 eQTLs and 3 eGenes in the disc degeneration-ALL multimorbidity including HLA-DQB2. While no population studies associate HLA-DQB2 with disc degeneration, its significant enrichment has been characterized by unique lower back pain DNA methylation signatures in human T cells (71). These findings demonstrate the utility of our approach in identifying multimorbidities of various complex diseases.

The whole blood GRN we constructed integrated common genetic variants, chromatin interactions and eQTL data. This network provides insights that expand beyond the identification of multimorbidity. Firstly, the spatially regulated protein-coding genes (median TPM >0.1) within the GRN are not statistically different in their expression levels (mean ± SD, 29.5 ± 330.5) from protein coding genes that are not

spatially regulated (mean ± SD, 203.1 ± 5597.3). Yet, protein coding genes that are regulated distally (in trans), are statistically more intolerant to LOF ($p < 0.0001$) when compared to protein coding genes that are expressed but not included in the GRN. Secondly, we identified a spike in eQTL-eGene interactions within the HLA region (6p21.3) in the whole blood GRN. This spike was specific to blood and was not present in the other tissues that were tested. This observation is consistent with the relative importance of this region in immune processes (e.g. regulation of inflammation, innate and adaptive immunity, antigen processing and presentation, autoimmunity, and the complement system (72)). The observation that tissue-specific spikes were observed at other chromosomal locations possibly highlights loci that are important for features specific to those tissue.

We have presented an approach that identifies multimorbid conditions of the disease of interest without the need for *a priori* selection of the interacting phenotypes. The molecular connections we have identified in this study represent high-value targets for subsequent investigation into asthma development, multimorbidity, and future therapeutic development.

Data Availability

The whole blood spatial gene regulatory network is available in figshare with the identifier 10.17608/k6.auckland.17067953.v1. The left ventricle spatial gene regulatory network is available in figshare with the identifier 10.17608/k6.auckland.18593432.v1. The adult brain cortex spatial gene regulatory network is available in figshare with the identifier 10.17608/k6.auckland.18592811.v1.

Code Availability

All data analysis was performed using R version 4.0.1 (73) and RStudio version 1.3.959 (74). Scripts used for data analysis are available at https://github.com/Genome3d/asthma_multimorbidities.

References

1. Mims,J.W. (2015) Asthma: definitions and pathophysiology. In *International forum of allergy & rhinology*. Wiley Online Library, Vol. 5, pp. S2–S6.
2. El-Husseini,Z.W., Gosens,R., Dekker,F. and Koppelman,G.H. (2020) The genetics of asthma and the promise of genomics-guided drug target discovery. Lancet Respir. Med., **8**, 1045–1056.
3. Ober,C. (2016) Asthma genetics in the post-GWAS era. Ann. Am. Thorac. Soc., **13**, S85–S90.
4. Vos,T., Lim,S.S., Abbafati,C., Abbas,K.M., Abbasi,M., Abbasifard,M., Abbasi-Kangevari,M., Abbastabar,H., Abd-Allah,F. and Abdelalim,A. (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet, **396**, 1204–1222.
5. Zhernakova,A., Van Diemen,C.C. and Wijmenga,C. (2009) Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat. Rev. Genet., **10**, 43–55.

6. Weatherburn,C.J., Guthrie,B., Mercer,S.W. and Morales,D.R. (2017) Comorbidities in adults with asthma: population-based cross-sectional analysis of 1.4 million adults in Scotland. Clin. Exp. Allergy, **47**, 1246–1252.
7. Bousquet,J., Anto,J., Auffray,C., Akdis,M., Cambon-Thomsen,A., Keil,T., Haahtela,T., Lambrecht,B.N., Postma,D.S. and Sunyer,J. (2011) MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. Allergy, **66**, 596–604.
8. Roberts,K.C., Rao,D.P., Bennett,T.L., Loukine,L. and Jayaraman,G.C. (2015) Prevalence and patterns of chronic disease multimorbidity and associated determinants in Canada. Heal. Promot. chronic Dis. Prev. Canada Res. policy Pract., **35**, 87.
9. Caughey,G.E., Vitry,A.I., Gilbert,A.L. and Roughead,E.E. (2008) Prevalence of comorbidity of chronic diseases in Australia. BMC Public Health, **8**, 1–13.
10. Haider,S., Simpson,A. and Custovic,A. Genetics of Asthma and Allergic Diseases. In. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–17.
11. Demenais,F., Margaritte-Jeannin,P., Barnes,K.C., Cookson,W.O.C., Altmüller,J., Ang,W., Barr,R.G., Beaty,T.H., Becker,A.B. and Beilby,J. (2018) Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat. Genet., **50**, 42–53.
12. Fadason,T., Farrow,S., Gokuladhas,S., Golovina,E., Nyaga,D., O'Sullivan,J.M. and Schierding,W. (2021) Assigning function to SNPs: Considerations when interpreting genetic variation. In *Seminars in Cell & Developmental Biology*. Elsevier.
13. Martin,P., McGovern,A., Orozco,G., Duffus,K., Yarwood,A., Schoenfelder,S., Cooper,N.J., Barton,A., Wallace,C. and Fraser,P. (2015) Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. Nat. Commun., **6**, 1–7.
14. Dekker,J. and Misteli,T. (2015) Long-range chromatin interactions. Cold Spring Harb. Perspect. Biol., **7**, a019356.
15. Ardlie,K.G., DeLuca,D.S., Segrè,A. V., Sullivan,T.J., Young,T.R., Gelfand,E.T., Trowbridge,C.A., Maller,J.B., Tukiainen,T., Lek,M., *et al.* (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-.).*, **348**.
16. Schierding,W., Farrow,S., Fadason,T., Graham,O.E.E., Pitcher,T.L., Qubisi,S., Davidson,A.J., Perry,J.K., Anderson,T.J. and Kennedy,M.A. (2020) Common variants Coregulate expression of GBA and modifier genes to delay Parkinson's disease onset. Mov. Disord., **35**, 1346–1356.
17. Fadason,T., Schierding,W., Kolbenev,N., Liu,J., Ingram,J.R. and O'Sullivan,J.M. (2020) Reconstructing the blood metabolome and genotype using long-range chromatin interactions. Metab. Open, **6**, 100035.
18. Goh,K.-I., Cusick,M.E., Valle,D., Childs,B., Vidal,M. and Barabási,A.-L. (2007) The human disease network. *Proc. Natl. Acad. Sci.*, **104**, 8685–8690.
19. Menche,J., Sharma,A., Kitsak,M., Ghiassian,S.D., Vidal,M., Loscalzo,J. and Barabási,A.-L. (2015) Uncovering disease-disease relationships through the incomplete interactome. Science (80-.)., **347**.

20. Sharma,A., Menche,J., Huang,C.C., Ort,T., Zhou,X., Kitsak,M., Sahni,N., Thibault,D., Voung,L. and Guo,F. (2015) A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. Hum. Mol. Genet., **24**, 3005–3020.
21. Sharma,A., Kitsak,M., Cho,M.H., Ameli,A., Zhou,X., Jiang,Z., Crapo,J.D., Beaty,T.H., Menche,J. and Bakke,P.S. (2018) Integration of molecular interactome and targeted interaction analysis to identify a COPD disease network module. Sci. Rep., **8**, 1–14.
22. Maiorino,E., Baek,S.H., Guo,F., Zhou,X., Kothari,P.H., Silverman,E.K., Barabási,A.-L., Weiss,S.T., Raby,B.A. and Sharma,A. (2020) Discovering the genes mediating the interactions between chronic respiratory diseases in the human interactome. Nat. Commun., **11**, 1–14.
23. Charoenngam,N., Ponvilawan,B., Wongtrakul,W. and Ungprasert,P. (2021) Patients with asthma have a higher risk of systemic lupus erythematosus: a systematic review and meta-analysis. Clin. Rheumatol., **40**.
24. Jiang,L., Sun,Y.Q., Langhammer,A., Brumpton,B.M., Chen,Y., Nilsen,T. Il, Leivseth,L., Wahl,S.G.F. and Mai,X.M. (2021) Asthma and asthma symptom control in relation to incidence of lung cancer in the HUNT study. Sci. Rep., **11**.
25. Kwon,J.H., Wi,C.-I., Seol,H.Y., Park,M., King,K., Ryu,E., Sohn,S., Liu,H. and Juhn,Y.J. (2021) Risk, Mechanisms and Implications of Asthma-Associated Infectious and Inflammatory Multimorbidities (AIMs) among Individuals With Asthma: a Systematic Review and a Case Study. Allergy. Asthma Immunol. Res., **13**, 697.
26. Lareau,C.A., Deweese,C.F., Adrianto,I., Lessard,C.J., Gaffney,P.M., Iannuzzi,M.C., Rybicki,B.A., Levin,A.M. and Montgomery,C.G. (2017) Polygenic risk assessment reveals pleiotropy between sarcoidosis and inflammatory disorders in the context of genetic ancestry. Genes Immun., **18**.
27. Fadason,T., Schierding,W., Lumley,T. and O'Sullivan,J.M. (2018) Chromatin interactions and expression quantitative trait loci reveal genetic drivers of multimorbidities. Nat. Commun., **9**.
28. Rao,S.S.P., Huntley,M.H., Durand,N.C., Stamenova,E.K., Bochkov,I.D., Robinson,J.T., Sanborn,A.L., Machol,I., Omer,A.D., Lander,E.S., *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell, **159**.
29. Szklarczyk,D., Gable,A.L., Lyon,D., Junge,A., Wyder,S., Huerta-Cepas,J., Simonovic,M., Doncheva,N.T., Morris,J.H., Bork,P., *et al.* (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res., **47**.
30. Durinck,S., Spellman,P.T., Birney,E. and Huber,W. (2009) Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. Nat. Protoc., **4**.
31. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res., **13**.
32. Benjamini,Y. and Hochberg,Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B, **57**.

33. Raudvere,U., Kolberg,L., Kuzmin,I., Arak,T., Adler,P., Peterson,H. and Vilo,J. (2019) G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res., **47**.
34. Jassal,B., Matthews,L., Viteri,G., Gong,C., Lorente,P., Fabregat,A., Sidiropoulos,K., Cook,J., Gillespie,M., Haw,R., *et al.* (2020) The reactome pathway knowledgebase. Nucleic Acids Res., **48**.
35. Karczewski,K.J., Francioli,L.C., Tiao,G., Cummings,B.B., Alföldi,J., Wang,Q., Collins,R.L., Laricchia,K.M., Ganna,A., Birnbaum,D.P., *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. Nature, **581**, 434–443.
36. Schmiedel,B.J., Singh,D., Madrigal,A., Valdovino-Gonzalez,A.G., White,B.M., Zapardiel-Gonzalo,J., Ha,B., Altay,G., Greenbaum,J.A., McVicker,G., *et al.* (2018) Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell, **175**.
37. Pividori,M., Schoettler,N., Nicolae,D.L., Ober,C. and Im,H.K. (2019) Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir. Med., **7**, 509–522.
38. Moffatt,M.F., Gut,I.G., Demenais,F., Strachan,D.P., Bouzigon,E., Heath,S., von Mutius,E., Farrall,M., Lathrop,M. and Cookson,W.O.C.M. (2010) A Large-Scale, Consortium-Based Genomewide Association Study of Asthma. N. Engl. J. Med., **363**, 1211–1221.
39. Almoguera,B., Vazquez,L., Mentch,F., Connolly,J., Pacheco,J.A., Sundaresan,A.S., Peissig,P.L., Linneman,J.G., McCarty,C.A. and Crosslin,D. (2017) Identification of four novel loci in asthma in European American and African American populations. Am. J. Respir. Crit. Care Med., **195**, 456–463.
40. von Mutius,E. and Smits,H.H. (2020) Primary prevention of asthma: from risk and protective factors to targeted strategies for prevention. Lancet, **396**.
41. Mills,M.C. and Rahal,C. (2019) A scientometric review of genome-wide association studies. Commun. Biol., **2**, 1–11.
42. Wong,C.K., Ho,C.Y., Ko,F.W.S., Chan,C.H.S., Ho,A.S.S., Hui,D.S.C. and Lam,C.W.K. (2001) Proinflammatory cytokines (IL-17, IL-6, IL-18 and IL-12) and Th cytokines (IFN-γ, IL-4, IL-10 and IL-13) in patients with allergic asthma. Clin. Exp. Immunol., **125**, 177–183.
43. Alam,R., York,J., Boyars,M., Stafford,S., Grant,J.A., Lee,J., Forsythe,P., Sim,T. and Ida,N. (1996) Increased MCP-1, RANTES, and MIP-1alpha in bronchoalveolar lavage fluid of allergic asthmatic patients. Am. J. Respir. Crit. Care Med., **153**, 1398–1404.
44. Watanabe,S., Suzukawa,M., Tashimo,H., Ohshima,N., Asari,I., Imoto,S., Kobayashi,N., Tohma,S., Nagase,T. and Ohta,K. (2021) High serum cytokine levels may predict the responsiveness of patients with severe asthma to benralizumab. J. Asthma, 10.1080/02770903.2021.1942039.
45. Kalayci,O., Birben,E., Tinari,N., Oguma,T., Iacobelli,S. and Lilly,C.M. (2004) Role of 90K protein in asthma and TH2-type cytokine expression. Ann. Allergy, Asthma Immunol., **93**, 485–492.
46. Wood,A.R., Perry,J.R.B., Tanaka,T., Hernandez,D.G., Zheng,H.F., Melzer,D., Gibbs,J.R., Nalls,M.A., Weedon,M.N., Spector,T.D., *et al.* (2013) Imputation of Variants from the 1000 Genomes Project

Modestly Improves Known Associations and Can Identify Low-frequency Variant - Phenotype Associations Undetected by HapMap Based Imputation. PLoS One, **8**.

47. Grunewald,J., Grutters,J.C., Arkema,E. V, Saketkoo,L.A., Moller,D.R. and Müller-Quernheim,J. (2019) Sarcoidosis. Nat. Rev. Dis. Prim., **5**, 1–22.
48. Starshinova,A.A., Malkova,A.M., Basantsova,N.Y., Zinchenko,Y.S., Kudryavtsev,I. V, Ershov,G.A., Soprun,L.A., Mayevskaya,V.A., Churilov,L.P. and Yablonskiy,P.K. (2020) Sarcoidosis as an Autoimmune Disease. Front. Immunol., **10**, 2933.
49. Chen,M.H., Li,C.T., Tsai,C.F., Lin,W.C., Chang,W.H., Chen,T.J., Pan,T.L., Su,T.P. and Bai,Y.M. (2014) Risk of Dementia Among Patients With Asthma: A Nationwide Longitudinal Study. J. Am. Med. Dir. Assoc., **15**.
50. Borgesius,N.Z., de Waard,M.C., van der Pluijm,I., Omrani,A., Zondag,G.C.M., van der Horst,G.T.J., Melton,D.W., Hoeijmakers,J.H.J., Jaarsma,D. and Elgersma,Y. (2011) Accelerated age-related cognitive decline and neurodegeneration, caused by deficient DNA repair. J. Neurosci., **31**, 12543–12553.
51. Dolle,M.E., Kuiper,R. V, Roodbergen,M., Robinson,J., Wijnhoven,S.W., Beems,R.B., de la Fonteyne,L., Niedernhofer,L.J., Hasty,P. and Vijg,J. (2011) Broad segmental progeroid changes in short-lived Ercc1 (-/Delta7) mice, Pathobiol. Aging Age Relat. Dis, **1**.
52. Jensen,H.L.B., Lillenes,M.S., Rabano,A., Günther,C.-C., Riaz,T., Kalayou,S.T., Ulstein,I.D., Bøhmer,T. and Tønjum,T. (2018) Expression of nucleotide excision repair in Alzheimer's disease is higher in brain tissue than in blood. Neurosci. Lett., **672**, 53–58.
53. Greten,F.R. and Grivennikov,S.I. (2019) Inflammation and cancer: triggers, mechanisms, and consequences. Immunity, **51**, 27–41.
54. Burgstaller,G., Oehrle,B., Gerckens,M., White,E.S., Schiller,H.B. and Eickelberg,O. (2017) The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. Eur. Respir. J., **50**.
55. Vignola,A.M., Kips,J. and Bousquet,J. (2000) Tissue remodeling as a feature of persistent asthma. J. Allergy Clin. Immunol., **105**, 1041–1053.
56. Simard,B., Turcotte,H., Marceau,P., Biron,S., Hould,F.S., Lebel,S., Marceau,S. and Boulet,L.-P. (2004) Asthma and Sleep Apnea in Patients with Morbid Obesity: Outcome after Bariatric Surgery. Obes. Surg., **14**, 1381–1388.
57. Akın,O., Sarı,E., Arslan,M., Yeşilkaya,E., Hacihamdioglu,B. and Yavuz,S.T. (2016) Association of wider neck circumference and asthma in obese children. Ann. Allergy, Asthma Immunol., **116**, 514–517.
58. Grossmann,N., Vakkasoglu,A.S., Hulpke,S., Abele,R., Gaudet,R. and Tampé,R. (2014) Mechanistic determinants of the directionality and energetics of active export by a heterodimeric ABC transporter. Nat. Commun., **5**, 1–10.
59. Williams,F.M.K., Bansal,A.T., van Meurs,J.B., Bell,J.T., Meulenbelt,I., Suri,P., Rivadeneira,F., Sambrook,P.N., Hofman,A. and Bierma-Zeinstra,S. (2013) Novel genetic variants associated with

lumbar disc degeneration in northern Europeans: a meta-analysis of 4600 subjects. Ann. Rheum. Dis., **72**, 1141–1148.

60. Liu,R., Chen,X. and Qi,J. (2018) Associations of TAP1 genetic polymorphisms with atopic diseases: asthma, rhinitis and dermatitis. Oncotarget, **9**, 1553.
61. Yang,G., Han,Y.Y., Forno,E., Yan,Q., Rosser,F., Chen,W. and Celedón,J.C. (2020) Glycated Hemoglobin A1c, Lung Function, and Hospitalizations Among Adults with Asthma. J. Allergy Clin. Immunol. Pract., **8**.
62. Matysiak,J., Klupczynska,A., Packi,K., Mackowiak-Jakubowska,A., Bręborowicz,A., Pawlicka,O., Olejniczak,K., Kokot,Z.J. and Matysiak,J. (2020) Alterations in Serum-Free Amino Acid Profiles in Childhood Asthma. Int. J. Environ. Res. Public Heal., **17**.
63. Panda,L., Gheware,A., Rehman,R., Yadav,M.K., Jayaraj,B.S., Madhunapantula,S. V, Mahesh,P.A., Ghosh,B., Agrawal,A. and Mabalirajan,U. (2017) Linoleic acid metabolite leads to steroid resistant asthma features partially through NF-κB. Sci. Rep., **7**, 9565.
64. Mabalirajan,U., Rehman,R., Ahmad,T., Kumar,S., Singh,S., Leishangthem,G.D., Aich,J., Kumar,M., Khanna,K., Singh,V.P., *et al.* (2013) Linoleic acid metabolite drives severe asthma by causing airway epithelial injury. Sci. Rep., **3**, 1349.
65. Zhou,Q., Fu,Y., Hu,L., Li,Q., Jin,M. and Jiang,E. (2018) Relationship of circulating chemerin and omentin levels with Th17 and Th9 cell immune responses in patients with asthma. J. Asthma, **55**, 579–587.
66. Chang,J.S., Tsai,Y.-W., Tsai,C.-R. and Wiemels,J.L. (2012) Allergy and risk of childhood acute lymphoblastic leukemia: a population-based and record-based study. Am. J. Epidemiol., **176**, 970–978.
67. Parker,B.R. (1997) Leukemia and lymphoma in childhood. Radiol. Clin. North Am., **35**, 1495–1516.
68. Garneau,A.P., Riopel,J. and Isenring,P. (2015) Acute Methotrexate-Induced Crystal Nephropathy. N. Engl. J. Med., **373**, 2691–2693.
69. Kishi,S., Yang,W., Boureau,B., Morand,S., Das,S., Chen,P., Cook,E.H., Rosner,G.L., Schuetz,E. and Pui,C.-H. (2004) Effects of prednisone and genetic polymorphisms on etoposide disposition in children with acute lymphoblastic leukemia. Blood, **103**, 67–72.
70. Livshits,G., Ermakov,S., Popham,M., MacGregor,A.J., Sambrook,P.N., Spector,T.D. and Williams,F.M.K. (2010) Evidence that bone mineral density plays a role in degenerative disc disease: The UK twin spine study. Ann. Rheum. Dis., **69**.
71. Grégoire,S., Cheishvili,D., Salmon-Divon,M., Dymov,S., Topham,L., Calderon,V., Shir,Y., Szyf,M. and Stone,L.S. (2021) Epigenetic signature of chronic low back pain in human T cells. Pain reports, **6**.
72. Shiina,T., Hosomichi,K., Inoko,H. and Kulski,J.K. (2009) The HLA genomic loci map: Expression, interaction, diversity and disease. J. Hum. Genet., **54**.
73. R core team (2021) R: A Language and Environment for Statistical Computing. R Found. Stat. Comput.

74. RStudio team (2020) RStudio: Integrated development environment for R.

Declarations

Acknowledgements

This work was funded by the Dines Family Charitable Trust (J.M.O., R.E.Z., and T.F.) and an HRC Explorer Grant (HRC 19/774; J.M.O. and T.F.). We thank the members of the Genomics and Systems Biology Group (Liggins Institute) for thoughtful discussions and suggestions. This work contains data from the Genotype-Tissue Expression (GTEx) Project, which was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Authors' contributions

R.E.Z. conceptualized, performed analysis, interpreted data, and wrote the manuscript. T.F. contributed to conceptualization, data analysis, data interpretation and manuscript revision. J.M.O. directed the study, contributed to data interpretation and data analysis, and co-wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and requests for materials should be addressed to J.M.O.

Figures

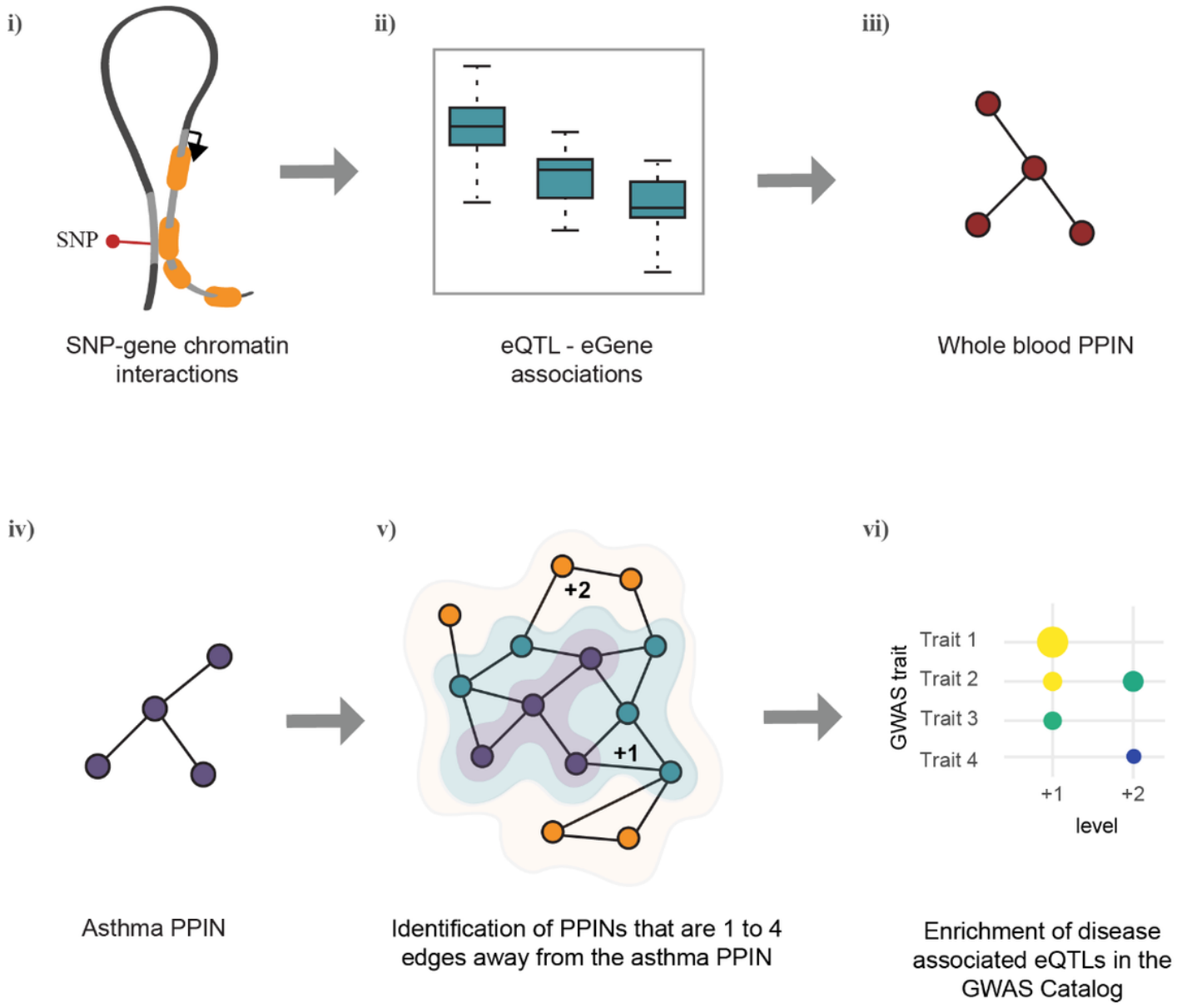

Figure 1

Study overview. i) Hi-C captured SNP-gene physical interactions in primary blood immune cells. ii) identification of SNP-Gene pairs having eQTL associations (GTEx v8). iii) Translation of the blood spatial gene regulatory network into a PPIN. iv) locating the asthma disease module within the blood PPIN. v) expanding four edges away from the asthma PPIN to identify neighboring PPIN in each of the four levels. vi) hypergeometric test to identify traits enriched in the four expanded PPINs.

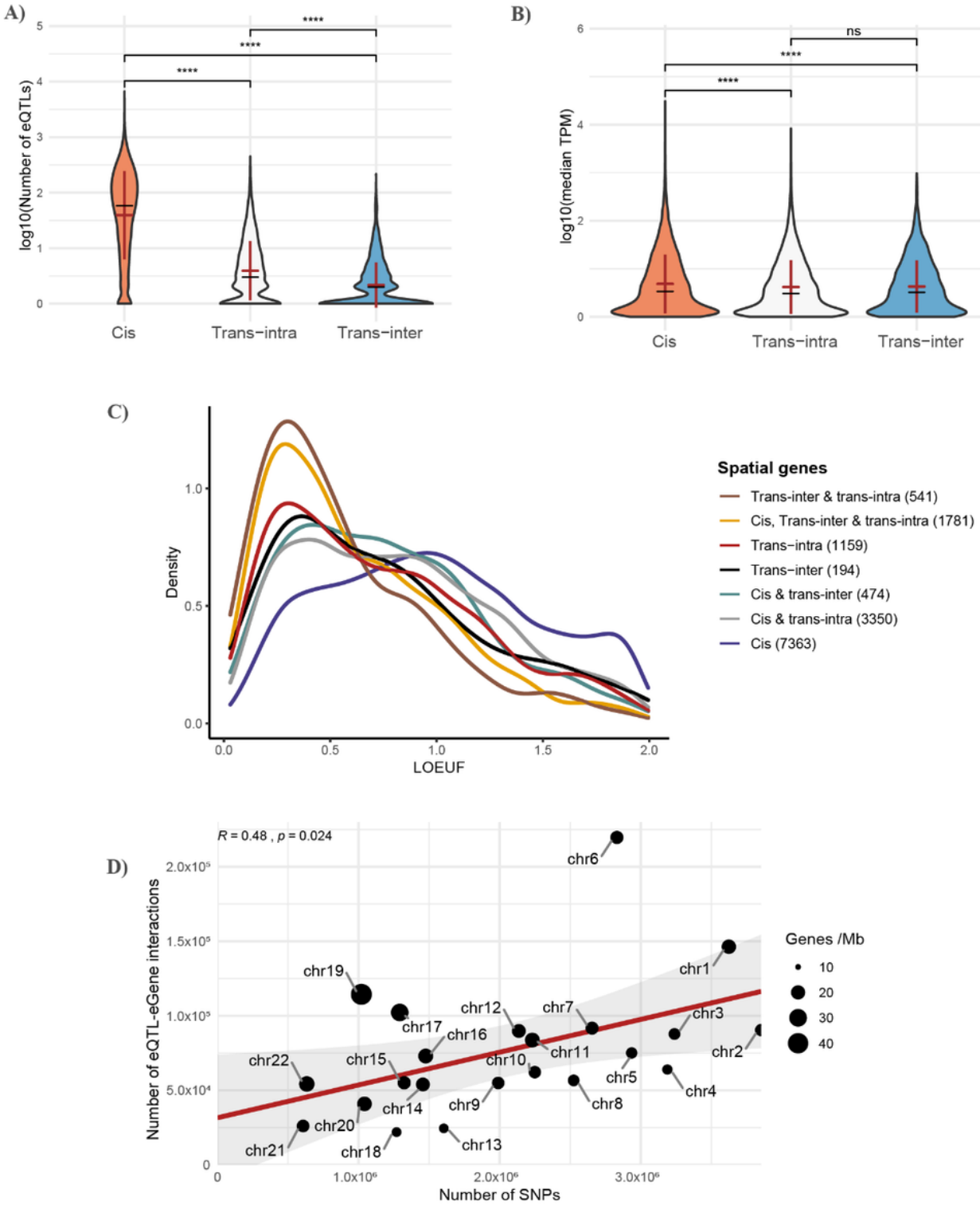

Figure 2

Characteristics of the blood GRN. Mean (red line), median (black line), and standard deviation of A) number of spatial eQTLs per gene grouped by interaction type and B) gene expression grouped by interaction type. C) Tolerance to loss of function of spatial eGenes grouped by interaction type. D) Correlation of blood GRN eQTLs and all variants genotyped from whole blood samples obtained from GTEx v8. NS, not significant, t test (****P <1 × 10−4).

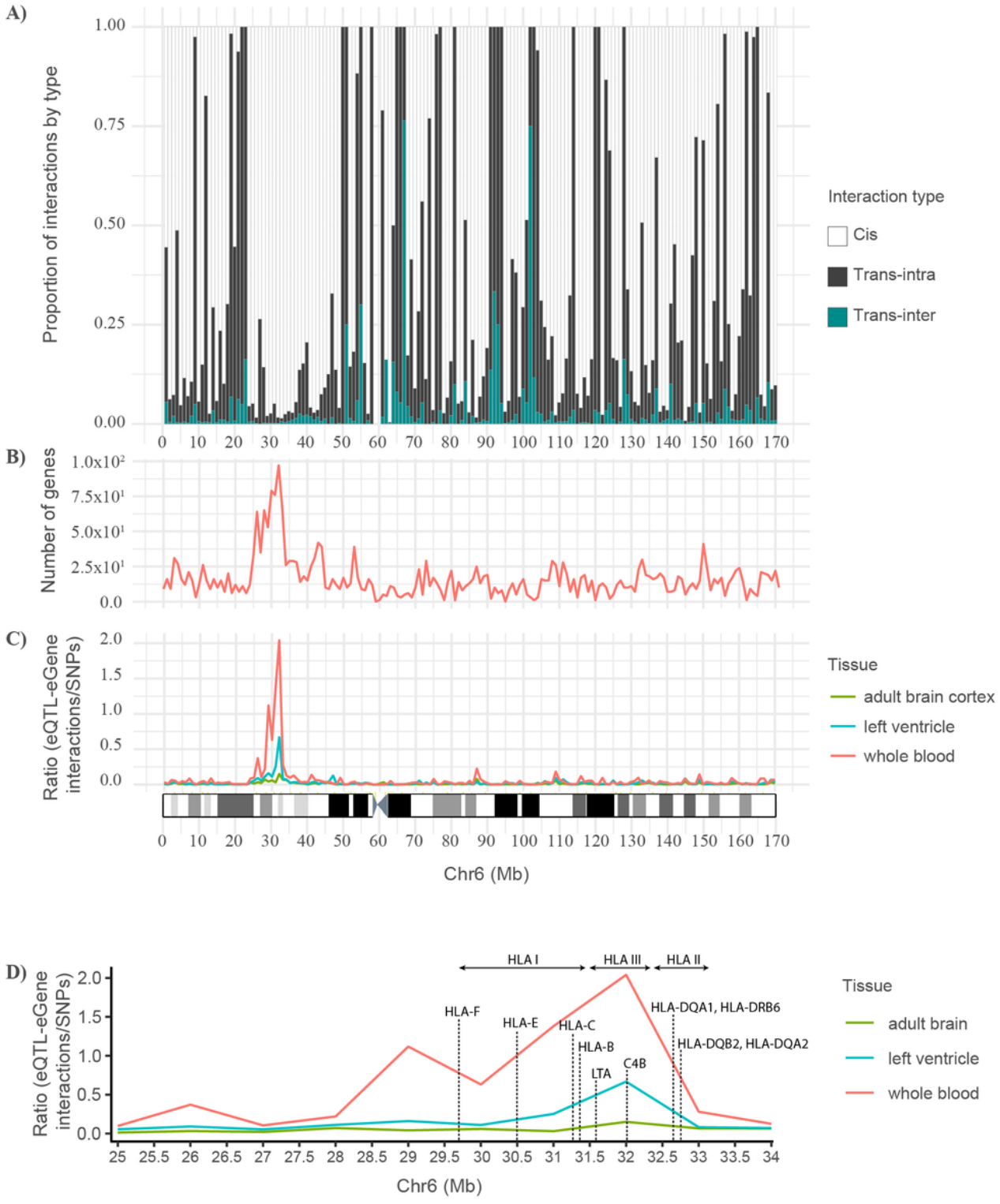

Figure 3

Tissue-specific spike of eQTL-eGene interactions in the HLA region of chromosome 6 in whole blood. A) Proportion of spatial eQTL-eGene interactions across chromosome 6 grouped by type B) Gene density across chromosome 6. C) Ratio of blood GRN eQTL-eGene interaction to SNPs genotyped in

chromosome 6; G-banding ideogram of chromosome 6 is shown below. D) magnified view of B showing the position of eGenes in the HLA region. A sliding window of 100 Mb was used for all plots.

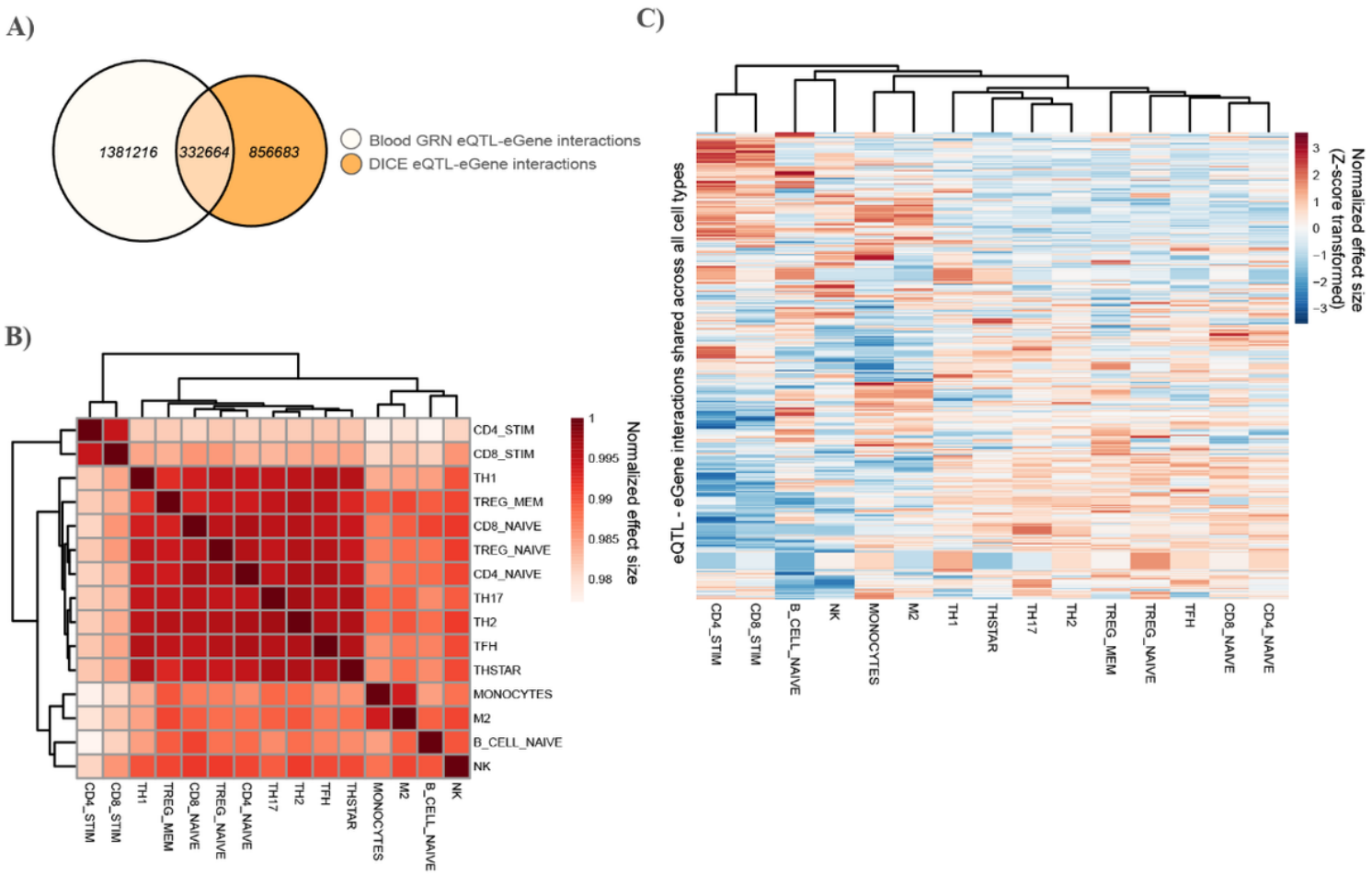

Figure 4

Spatial eQTL-eGene interactions shared across immune cell types show immune cell clusters having similar spatial gene regulatory pathways. A) Venn diagram of blood GRN eQTL-eGene interactions and all DICE eQTL-eGene interactions. B) Correlation matrix of eQTL-eGene interactions shared between DICE and the blood GRN. Stronger correlations are represented with darker shades of red. C) Heatmap of Z-score-transformed normalized effect size of eQTL-eGene interactions shared between DICE and the blood GRN showing spatial blood GRN eQTLs as rows and DICE cell types as columns. In both B and C, only DICE cell types sharing eQTL-eGene interactions were included.

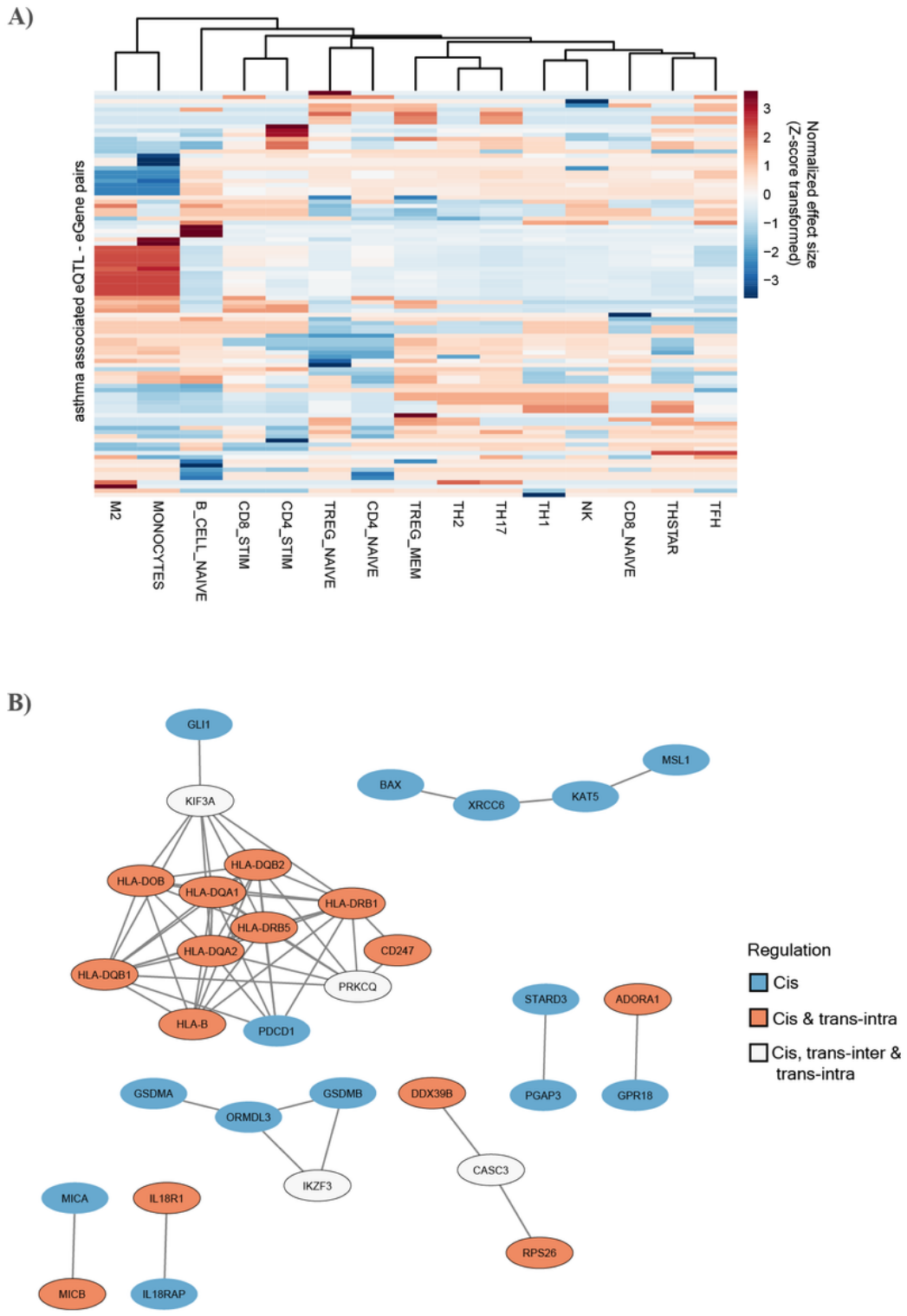

Figure 5

Immune cell types sharing similar asthma GRNs suggest their potential shared regulatory roles in the etiology of asthma. A) Unsupervised clustering of z-score transformed normalized effect size of all asthma-associated spatial eQTL-eGene interactions (rows) and DICE immune cell types sharing eQTL-gene interactions (columns). B) Asthma protein-protein interaction network (only proteins having edges

are illustrated). Node color corresponds to the type of regulation. Visualization was done using Cytoscape (version 3.8.2).

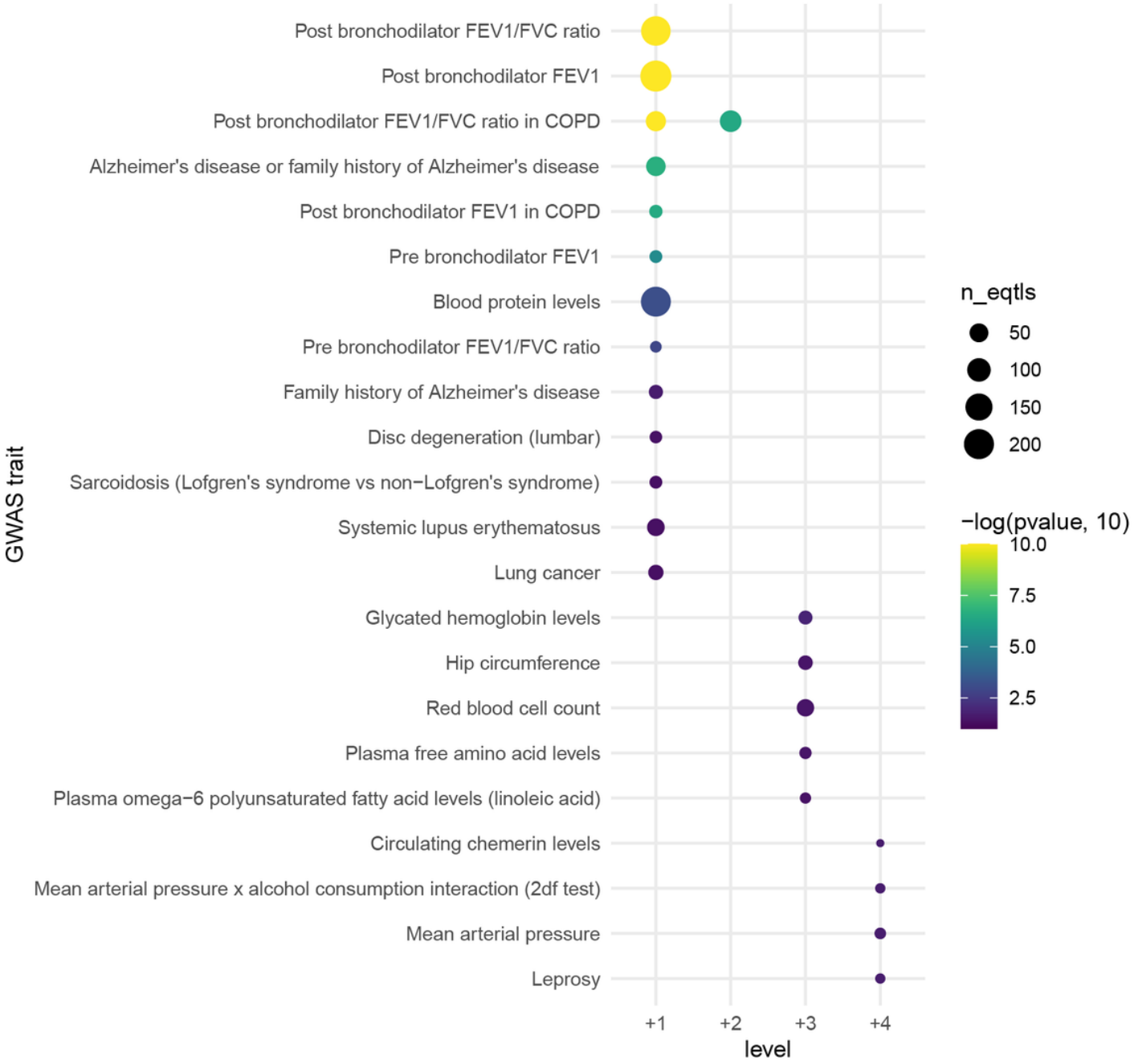

Figure 6

Blood GRN *de novo* identifies conditions that are multimorbid with asthma. GWAS traits enriched in each of the four expanded PPIN neighbors of the asthma disease module. Circle size indicates the number of eQTLs, and circle color indicates statistical significance. FEV1; forced expiratory volume in the first second, FVC; Forced vital capacity, COPD; chronic obstructive pulmonary disease.

Supplementary Files

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

- asthmamultimorbiditiessuppldata3.xlsx
- asthmamultimorbiditiessupplfigs.docx
- asthmamultimorbiditiessuppldata10.xlsx
- asthmamultimorbiditiessuppldata5.xlsx
- asthmamultimorbiditiessuppldata6.xlsx
- asthmamultimorbiditiessuppldata2.xlsx
- asthmamultimorbiditiessuppldata9.xlsx
- asthmamultimorbiditiessuppldata8.xlsx
- asthmamultimorbiditiessuppldata7.xlsx
- asthmamultimorbiditiessuppldata1.xlsx
- asthmamultimorbiditiessuppldata4.xlsx
- NCOMMS2203625RS.pdf