Implementation of an integrated network analysis strategy for multi-omics data: Application to GWAS data of allergic diseases and EWAS data of IgE levels

Eranti Pradeep, Vernet Raphaël, Linhard Christophe, Mohamdi Hamida, EGEA, EVADA and RESET-AID consortia, Bouzigon Emmanuelle, Demenais Florence

Group of Genomic Epidemiology and Multifactorial Diseases, Université Paris Cité, Inserm, UMRS-1124, F-75006 Paris, France.

1 Introduction

Allergic diseases (AD), including asthma, eczema and hay fever, are multifactorial diseases stemming from the effects and interactions of multiple genetic and environmental factors. AD is characterized by increased levels of total and specific Immunoglobulin E (IgE) levels and positive responses to skin prick tests to allergens. Total serum IgE is a central mediator of allergic inflammation. It is likely that shared biological mechanisms are underlying IgE production and the risk of AD. While the number of genetic variants associated with AD risk has been increasing in recent studies [1–5], GWASs of IgE levels have identified a small number of genetic variants which account for only 1-2% of the variation in serum IgE. Epigenetic mechanisms, such as CpG methylation and microRNAs, regulate the expression of genes that may play a role in IgE production. Recent epigenome-wide association studies of total IgE have revealed CpG sites in promoter regions of genes accounting for variation in total serum IgE levels tenfold higher than that derived from large genome-wide association studies [6]. It can be hypothesized that integrating results (summary statistics) from GWAS of AD and EWAS of IgE levels can provide a more comprehensive understanding of the molecular mechanisms underlying IgE regulation and the risk of AD.

To appreciate the success of genome-wide approaches and also to include the factors with small marginal effects in the downstream analysis, methods integrating the biological knowledge and genome-wide study findings have been proposed. Among these methods, network analysis that integrates the genome-wide results with a biological network to identify sub-networks enriched in disease association signals has become more broadly used and shown to be successful [7–11]. It can be argued that network analysis provides a promising approach to discovering functionally related genes acting jointly with small marginal effects.

The goal of this study was to identify gene modules enriched in allergy-associated signals through network analysis of outcomes of genome-wide association studies of AD with genetic variants (GWAS) and of IgE levels with DNA methylation profiles (EWAS). Our specific aims were to explore the relationships between AD-GWAS and IgE-EWAS derived gene modules through overlap and connectivity analyses of the module genes, and to identify joint AD-GWAS and IgE-EWAS modules including genes with direct connections between them. This analysis was followed by the functional enrichment analysis of the AD-GWAS and IgE-EWAS joint modules to better understand the complex mechanisms underlying the risk of allergic diseases and IgE production.

2 Datasets

We used three genome-wide datasets in this study: (1) results from the meta-analysis of 12 AD GWASs [3], (2,3) results from two epigenome-wide association studies of IgE levels in asthmatics and non-asthmatics performed by our group in the French Epidemiological study on the Genetics and Environment of Asthma (EGEA) study (unpublished results).

2.1 Allergy GWAS dataset

To identify a gene module associated with AD, we used an AD GWAS dataset that contained the meta-analysis results of 12 studies, described elsewhere [3]. The dataset included 8,183,608 autosomal SNPs from 96,794 cases (who reported suffering from asthma and/or eczema and/or hay fever) and 145,775 controls (who never reported suffering from any of these three disease conditions), all of the European ancestry. The association between each SNP and AD was performed using an additive genetic model, and the results were meta-analyzed using a fixed-effects model. In total, 8,183,608 SNPs with meta-analysis results (P-values of association with allergic diseases) were used as input for the network analysis.

2.2 IgE EWAS datasets

In order to identify gene modules associated with IgE levels, we used two EWAS datasets generated in asthmatics and non-asthmatics, respectively, from the EGEA study. DNA Methylation profiling was performed using MethylC-Capture Sequencing for blood samples from 713 adults (343 asthmatics + 370 non-asthmatics). This approach based on bisulfite sequencing allows targeting about ~5M CpGs near regulatory elements of immune cells, such as enhancers and promoters, which are not completely covered within existing Illumina arrays. After applying several quality control (QC) procedures and filtering steps [12], including our own QC procedure, 3.16 million (asthmatics) and 4.28 million (non-asthmatics) CpGs were available for analysis. The usage of different sequencers can mainly explain the difference in the number of CpGs available in asthmatics and non-asthmatics (HiSeq in the former and NovaSeq in the latter), which satisfied different QC criteria. Epigenome-wide association analysis between methylation profiles (distribution of the number of methylated reads among the total number of reads for each subject) and IgE levels was conducted using a Binomial Mixed Model implemented in the MACAU software [13]. Analyses were performed separately in asthmatic and non-asthmatic samples because the percent methylation profiles from the two groups showed poor clustering. This also allowed to compare the results between the two groups. In total, 3,159,712 (asthmatics) and 4,279,850 (non-asthmatics) CpGs with summary results were used as input for network analysis.

3 Methods

Recently, we proposed an analysis strategy to integrate multiple omics datasets with a publicly available protein-protein interaction network to identify gene modules associated with respective phenotypes of interest and to characterize the genes and pathways shared between the identified modules [14]. This study employs an updated version of this proposed strategy to analyze the outcomes of AD-GWAS and IgE-EWAS and, in particular, to explore the interconnections between AD-GWAS and IgE-EWAS module genes more extensively. This section will discuss the various steps involved in the analysis strategy.

3.1 Computing gene-level statistics from SNP-level or CpG-level statistics

The first step in the analysis strategy is to map SNPs and CpGs to genes, followed by the computation of gene-level statistics from SNP-level or CpG-level statistics. With respect to GWAS data, we chose to focus on SNPs affecting gene expression (cis-eQTLs). SNPs were assigned to genes using information from a cis-eQTL database (version 2018-06-27[1]), which included 126 publicly available cis-eQTL datasets from various tissues relevant to asthma, and the EUGENE procedure [15]. Gene P-values were calculated using the fastBAT method, as proposed by Bakshi et al. [16] and made available within EUGENE. This method calculates the gene statistic as a sum of test statistics (chi-squares) of the association of individual SNPs (here, cis-eQTLs or their proxies) with disease using the Satterthwaite's approximation to correct for correlation between SNPs. The SNP pairwise correlation was estimated using a reference dataset (here, 1000 Genomes European individuals). Regarding EWAS data, CpG sites were assigned to genes using the GREAT tool [17] if and only if the CpGs were located within 1Mb on either side of a genomic region (nearest gene option). We used the best CpG P-value among all CpGs assigned to the gene and computed corrected P-values using the Circular Genomic Permutation (CGP) strategy (fastCGP [9]) to take into account correlations between CpGs and gene length.

3.2 Construction of a scored Protein-Protein Interaction Network

Protein-protein interaction network from the STRING database [18] contains information on gene/protein relationships, including direct (physical) and indirect (functional) interactions. For Homo sapiens, the network contains 19,566 genes/proteins and 11,759,454 protein-protein interactions (PPI). The human PPI network (GeneNet) from the STRING database was downloaded (https://string-db.org/; version 11), and the gene P-values, calculated above, were converted into gene scores using the inverse normal distribution ($z = \phi^{-1}(1-P)$). Then, the calculated gene scores were overlaid on to the GeneNet to construct a scored network, e.g., AD-GWAS GeneNet is the scored network constructed by overlaying the gene scores resulting from the AD-GWAS study.

3.3 Identification of a gene module enriched with disease/trait association signals

The SigMod method [10] takes the scored GeneNet as input to identify a strongly interconnected module enriched with high association signals from the background scored network (using AD GWAS & IgE EWAS data, respectively). SigMod treats the task of identifying the list of genes present in the interconnected module as a selection problem by solving the maximization problem:

$$argmax_{u_1,u_2,\cdots,u_n}\, g(u_1,u_2,\cdots,u_n,\lambda,\eta), \qquad \begin{matrix}(u_p=1)\text{ a gene is selected}\\(u_p=0)\text{ or not selected}\end{matrix} \qquad (1)$$

[1] The database is available here https://genepi.qimr.edu.au/staff/manuelf/eugene/eqtl_info/v20180627/cis/r2-0.05/db-files/eqtl.20180627.ASTHMA.cis

where the objective function is defined as:

$$g(u_1, u_2, \cdots, u_n, \lambda, \eta) = \underbrace{\sum_i z_i u_i}_{Association} + \underbrace{\lambda \sum_{i \neq j} A_{ij} u_i u_j}_{Interconnectivity} - \underbrace{\eta \sum_i I(u_i = 1)}_{Sparsity} \quad \lambda \geq 0, \quad \eta \geq 0 \quad (2)$$

In the objective function g, z_i is the score of the i^{th}gene, $A_{ij}$is the interaction weight between i^{th} and j^{th} genes (A_{ij}= 0 if there is no interaction), u_i is an indicator variable indicating whether a gene is selected ($u_i = 1$) or not ($u_i = 0$). λ is the interconnection tuning parameter, and a big λ value will lead to selecting a strongly interconnected module. η is the sparsity tuning parameter, and a big η value will lead to select less genes. For fixed values of λ and η, solving the optimization problem defined by Equation (1) will determine whether a gene is selected or not, i.e. the value of each u_i. This maximization problem can be solved exactly using a graph-cut algorithm [7,10]. For further details about the steps involved in the updated version of the SigMod method, please refer to [14].

Assessment of the association of the selected gene module with the disease/trait

We assessed whether the final module selected by SigMod was significantly associated with: (i) the disease (AD) by estimating the null distribution of the module score (sum of scores of the genes belonging to the selected module) through 100,000 permutations of the SNP-level P-values and computation of the gene-level statistics using the fastBAT method (ii) the trait (IgE) by estimating the null distribution of the module scores by permuting the CpG-level P-values 100,000 times and computing, for each CGP sample, the gene-level P-values by the fastCGP method. The module score was calculated at each permutation using these recomputed gene P-values. Each observed GWAS or EWAS-derived module score was compared with the module scores obtained from the respective permutation samples to get an empirical p-value.

3.4 Assessment of the overlap between the two gene modules derived from the AD-GWAS and IgE-EWAS data

The overlap between the AD-GWAS module and each of the IgE-EWAS modules (asthmatics and non-asthmatics) identified by SigMod was estimated using the overlap coefficient $C = n_{shared}/min(n_{m_1}, n_{m_2})$ and the Jaccard Index $J = n_{shared}/(n_{m_1} + n_{m_2} - n_{shared})$, where n_{shared} denotes the number of genes shared between two modules m_1 and m_2, $n_{m_1}$and n_{m_2} denotes the number of genes in the modules m_1 and $m_2$respectively. These similarity measures lie in the range [0,1]; where $C, J = 0$ means there is no overlap between two modules, $C = 1$ signifies that one module is a complete subset of the other module, whereas $J = 1$ indicates two identical gene modules. The statistical significance of the overlap between modules was evaluated by the Hypergeometric test (P_{HYPER}).

3.5 Network connectivity between the gene modules derived from the AD-GWAS and IgE-EWAS data

We explored the connectivity between AD-GWAS and IgE-EWAS derived gene modules and identified the genes having direct connections (i.e. one edge or none for shared genes) between the AD-GWAS module and each of the two IgE-EWAS modules (asthmatics and non-asthmatics) on the GeneNet. We constructed two joint modules: one derived from AD-GWAS and IgE-EWAS in asthmatics and another one derived from AD-GWAS and IgE-EWAS in non-asthmatics. Each of these joint modules contained the directly connected genes as nodes and high-confidence interactions on the GeneNet as edges. We assessed the significance of the proportions of genes having a direct connection between the AD-GWAS module and either one of the two IgE-EWAS modules by randomly sampling from the genes on AD-GWAS or IgE-EWAS scored GeneNets the same number of genes as observed in each module and by calculating the proportion of genes involved in direct interactions. This procedure was repeated 100,000 times to get empirical P-values.

3.6 Biological interpretation of the joint AD-GWAS/IgE-EWAS gene modules

To explore the biological significance of the genes belonging to the joint AD-GWAS/IgE-EWAS modules, we conducted enrichment analysis (Over Representation Analysis) by analyzing the list of genes in the context of prior biological knowledge. We used clusterProfiler tools [19] by interrogating Gene Ontology (GO), KEGG and REACTOME databases. The significance of enriched terms/pathways was obtained by using the hypergeometric test and the Benjamini-Hochberg procedure to correct for multiple testing [20]. The resulting corrected P-value will be designated as P_{BH}. Based on the assumption that similar terms/pathways do not bring additional knowledge, we selected one representative GO term from a group of overlapping redundant GO terms using semantic similarity analysis (*simplify function* in clusterProfiler). Similarly, REACTOME pathways were grouped through pairwise similarity analysis based on the Jaccard index. The representative term/pathway of a group was the one with the smallest P_{BH} value [21].

4 Results

The different steps of the proposed analysis strategy and the main outcomes at each step are presented in Figure 1.

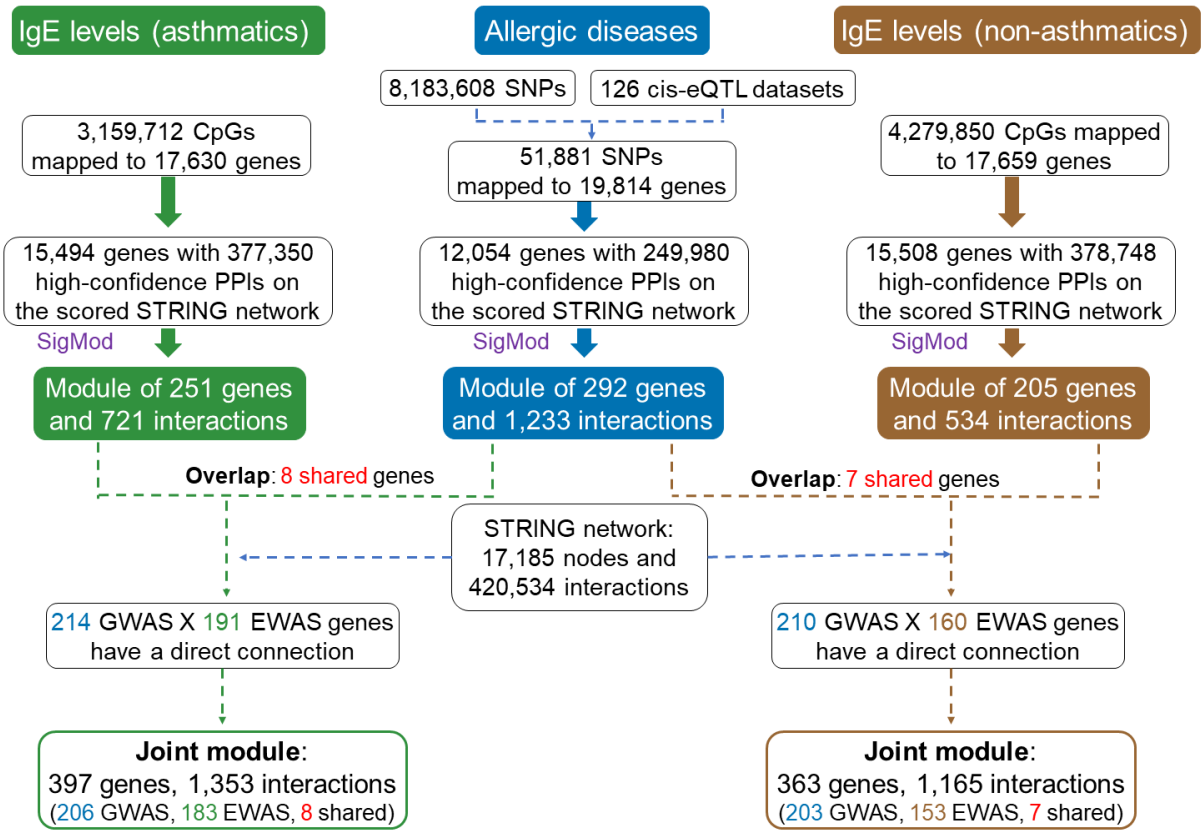

Figure 1: Summary of the steps involved in the analysis strategy to identify joint gene modules based on the SigMod method, along with the number of SNPs/CpGs/Genes present at each step.

4.1 Identification of gene modules enriched with disease/trait association signals from GWAS and EWAS data

AD-GWAS derived module: The source data included the summary statistics for 8,183,608 autosomal SNPs from the meta-analysis of 12 AD-GWASs. After proceeding with the EUGENE-based analysis of cis-eQTLs (i.e., significantly associated with gene expression at a P-value threshold of 8.9×10^{-10} and being in low linkage disequilibrium) or their proxies mapping to GWAS summary data, there were 51,881 SNPs mapped to at least one of 19,814 genes. When the gene-level P-values, computed using fastBAT and converted to gene scores, were overlaid onto the GeneNet of high-confidence interactions, we obtained the AD-GWAS scored GeneNet of 12,054 genes and 249,980 interactions among them. The use of SigMod allowed us to identify a strongly interconnected gene module containing 292 genes and 1,233 interactions, as shown in Figure S1. All 292 genes were associated with AD at $P<0.01$ (Figure 2).

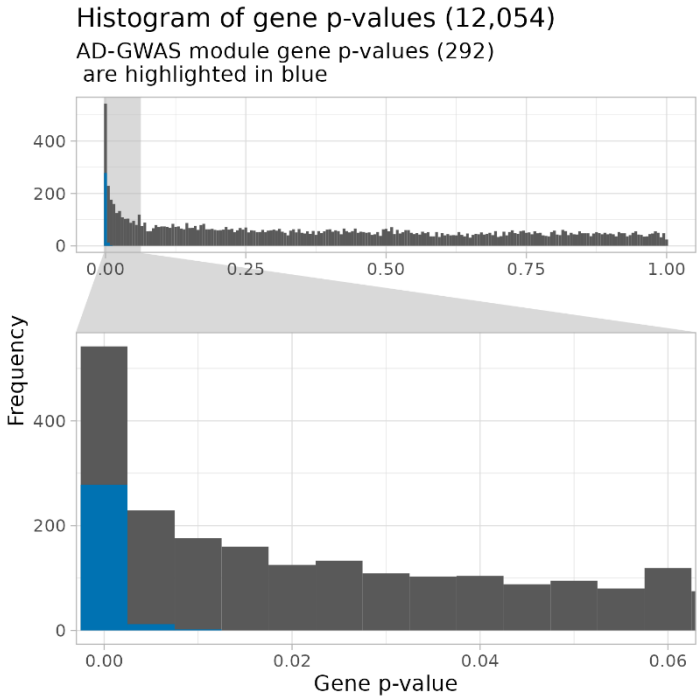

Figure 2: Histogram of the gene P-values (n=292) of the selected gene module from AD GWAS (in blue) and the gene P-values of all genes (12,054) present in the background network (in grey). All 292 genes are associated with AD at P <0.01. The bottom figure is a zoomed view of the top figure.

IgE-EWAS derived modules: We used 3,159,712 (asthmatics) and 4,279,850 (non-asthmatics) CpGs with summary results as input for the network analysis. After mapping these CpG sites to genes using GREAT and computing gene-level P-values using fastCGP, the gene scores derived from gene P-values were overlaid onto the GeneNet of high-confidence interactions to generate the IgE-EWAS scored GeneNet of 15,494 genes and 377,350 interactions for asthmatics, and 15,508 genes and 378,748 interactions for non-asthmatics. After applying SigMod, we obtained a gene module of 251 genes and 721 interactions in asthmatics (Figure S3) and a gene module of 205 genes and 534 interactions in non-asthmatics (Figure S4). More than 99% of the genes in each of the two IgE-EWAS derived modules were nominally associated with IgE levels (Figure 3 for asthmatics and Figure 4 for non-asthmatics).

We evaluated the association of each of these three gene modules with their respective trait (AD/IgE levels) by applying a permutation strategy. We calculated the module scores for 100,000 permutations and observed that none of these scores was higher than the observed respective module scores. The distribution of these scores can be seen in Figure S2, Figure S5, and Figure S6.

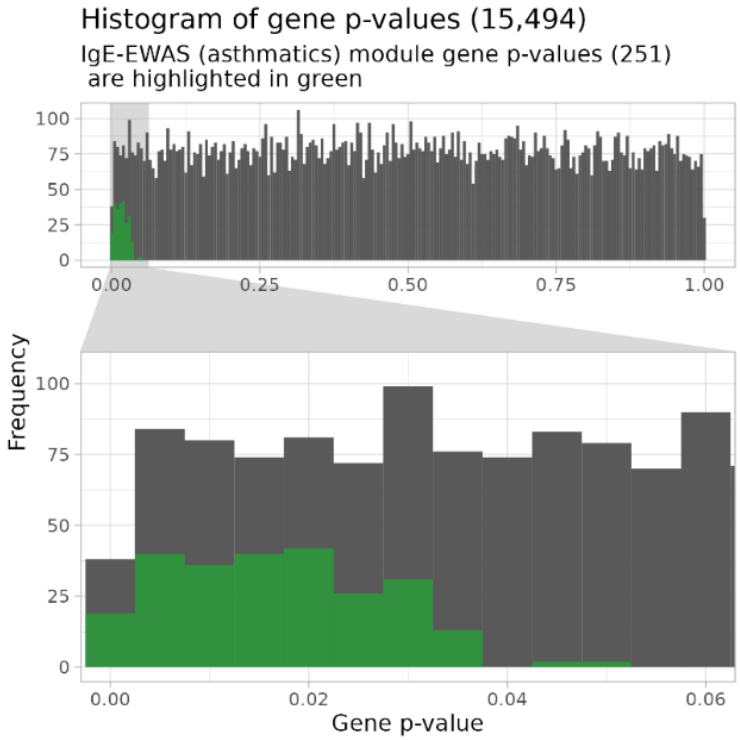

Figure 3: Histogram of the gene P-values (n=251) of the selected gene module from IgE EWAS in asthmatics (in green) and the gene P-values of all the genes (15,494) present in the background network (in grey). > 99% of the genes are nominally associated (P <= 0.05) with IgE levels. The bottom figure is a zoomed view of the top figure.

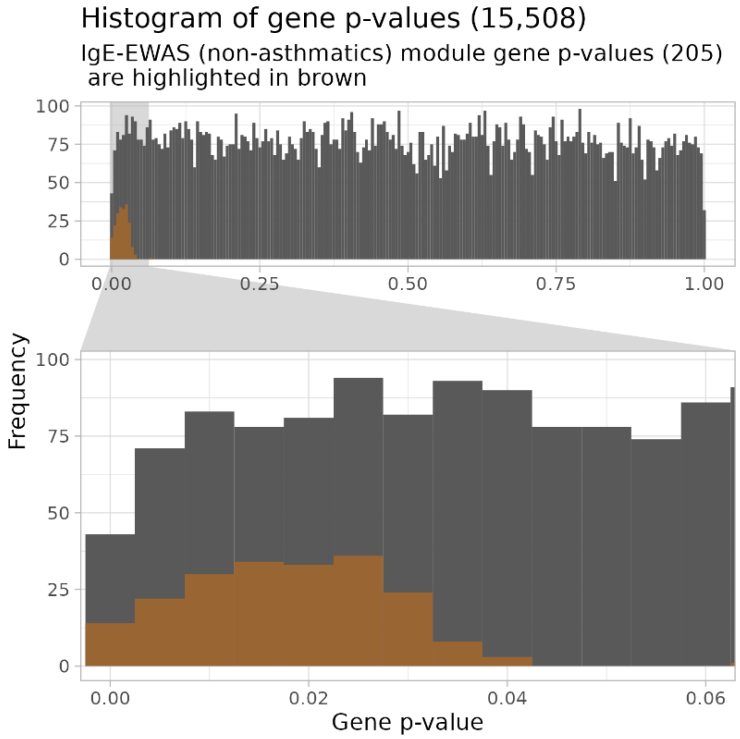

Figure 4: Histogram of the gene P-values (n=205) of the selected gene module from IgE EWAS in non-asthmatics (in brown) and the gene P-values of all the genes (15,508) present in the background network (in grey). > 99% of the genes are nominally associated (P <= 0.05) with IgE levels. The bottom figure is a zoomed view of the top figure.

4.2 Overlap of and connectivity between AD-GWAS and IgE-EWAS derived gene modules

We looked at the overlap between AD-GWAS and IgE-EWAS derived modules, and found that eight and seven genes were shared between the AD-GWAS module and either the IgE-EWAS module in asthmatics (P-value for overlap = 0.043) or IgE-EWAS module in non-asthmatics (P-value for overlap = 0.036).

A proportion of 73% of AD-GWAS module genes directly interacted with 76% of the IgE-EWAS module genes in asthmatics ($P<10^{-5}$ based on 100K permutations). We constructed the AD-GWAS/IgE-EWAS joint module in asthmatics by including all these interacting genes: 397 genes and 1,353 direct interactions (Figure 5). Similarly, 72% of the genes from the AD-GWAS module had direct interactions with 78% of IgE-EWAS module genes in non-asthmatics ($P<10^{-5}$ based on 100,000 permutations). This AD-GWAS/IgE-EWAS joint module in non-asthmatics included 1,165 direct interactions among 363 genes (Figure 6). We computed the average distance between AD-GWAS module genes and either of the two IgE-EWAS module genes for 100K samplings, and observed that none of these average distances is smaller than the observed respective average distance. This observation hints that the trait-associated genes are localized in a closed neighborhood.

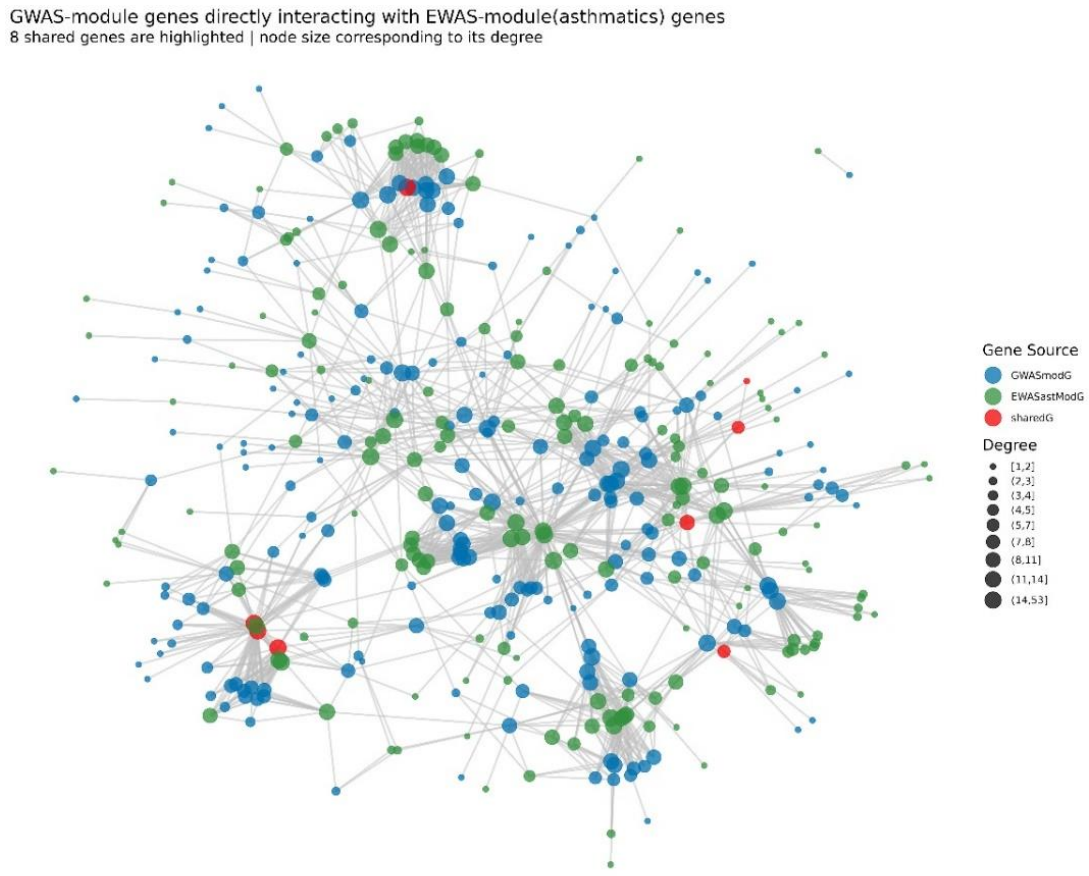

Figure 5: Joint module built from the direct connections between AD-GWAS and IgE-EWAS (asthmatics) gene modules.

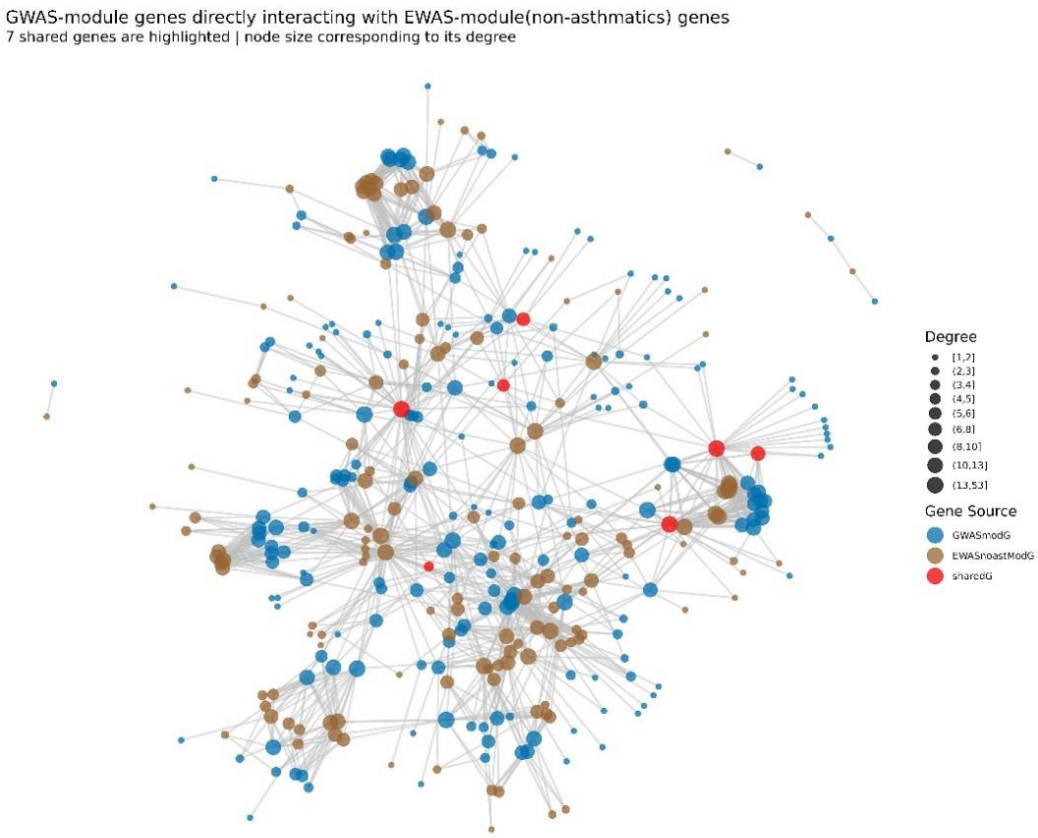

Figure 6: Joint module built from the direct connections between AD-GWAS and IgE-EWAS (non-asthmatics) gene modules

4.3 Biological Interpretation of the two joint modules

We performed enrichment analysis for genes belonging to each of the two joint modules using Gene Ontology, KEGG, and REACTOME databases through the clusterProfiler tools. After grouping redundant GO terms, there were 841 (asthmatics) and 796 (non-asthmatics) GO-BP categories, of which 84 (asthmatics) and 80 (non-asthmatics) GOs belonging to the top 10% of P_{BH} distribution were significantly enriched ($P_{BH} < 10^{-3}$) by genes from either one of the two joint modules. The top-ranked GO-BP categories were Antigen processing and presentation, Regulation of leukocyte T-cell adhesion, Regulation of T-cell activation, Adaptive immune response, and MHC class II protein assembly ($P_{BH} < 5 \times 10^{-18}$) for both joint modules. Among the top 5% significantly enriched GO-BP terms ($P_{BH} < 10^{-5}$), more than half were similar in the two groups of asthmatics and non-asthmatics.

Regarding the KEGG pathways (274 and 257 for the joint modules in asthmatics and non-asthmatics), the top 10% pathways were highly significantly enriched ($P_{BH} < 10^{-5}$). The vast majority of these pathways were common to both asthmatic and non-asthmatic groups, and included Allograft rejection, Antigen processing and presentation, and pathways related to auto-immune diseases, asthma and infectious diseases. With respect to REACTOME pathways (330 and 332 grouped pathways for the two respective joint modules), the top 10% showed strong significant enrichment ($P_{BH} < 2.5 \times 10^{-5}$) for both joint modules. A high proportion of

pathways were common to the two groups of asthmatics and non-asthmatics. The top pathways shared by the two groups included Interferon gamma signaling, MHC class I and MHC class II antigen presentation, PD-1 signaling, Co-stimulation by the CD20 family, and Signaling by interleukins. Among the top 10% pathways, only eight (24%) were specific to the asthmatic group, and eleven (33%) were specific to non-asthmatics (when looking at all pathways having $P_{BH} \leq 0.05$). The asthmatic-specific REACTOME pathways included the ER-phagosome pathway, RUNX1 regulates transcription of genes involved in differentiation of HSCs (hematopoietic stem cells), MyD88-MAL (TIRAP) cascade initiated on plasma membrane, Beta-catenin independent WNT signaling while the non-asthmatic specific pathways included Downstream TCR signaling, TCF dependent signaling in response to WNT, Estrogen- dependent gene expression, Chromatin modifying enzymes.

5 Conclusion

This study identified three gene modules enriched in association signals from AD-GWAS and IgE-EWAS conducted in asthmatics and non-asthmatics. All the interacting module genes were nominally significant. However, most of them did not reach the genome-wide significant level and, thus, would have been missed by individual gene-based analysis, reiterating the importance of network analysis in the pursuit of selecting candidate genes for further analysis. Interestingly, a significantly high proportion of AD-GWAS module genes directly interacted with genes from each IgE-EWAS module (more than 70%, $P<10^{-5}$), showing strong relationships between GWAS-derived genes associated with AD and EWAS-derived genes associated with IgE levels. Enrichment analysis of the joint AD-GWAS and IgE-EWAS modules indicated that a high proportion of significantly enriched biological pathways are common to the two groups of asthmatics and non-asthmatics. At the same time, a few are specific to each group. Altogether, this study provides better insight into the biological mechanisms underlying the risk of allergic diseases and the regulation of IgE production and opens new routes for further molecular investigations of our findings.

Funding and Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813533. This project was also supported by PHRC-CL-20131610 EVADA and ANR-15-EPIG-004-05-RESET-AID grants. We also thank the EGEA cooperative group (https://egeanet.vjf.inserm.fr/index.php/fr/groupe-egea), the EVADA and RESET-AID consortium members for their contribution to this work.

References

1. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmüller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet. 2018;50: 42–53.
2. Ferreira MA, Mathur R, Vonk JM, Szwajda A, Brumpton B, Granell R, et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. Am J Hum Genet. 2019;104: 665–684.
3. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet. 2017;49: 1752–1757.
4. Paternoster L, Standl M, Waage J, Baurecht H, Hotze M, Strachan DP, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat Genet. 2015;47: 1449–1456.
5. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. Nat Genet. 2018;50: 1072–1080.
6. Liang L, Willis-Owen SAG, Laprise C, Wong KCC, Davies GA, Hudson TJ, et al. An epigenome-wide association study of total serum immunoglobulin E concentration. Nature. 2015;520: 670–674.
7. Azencott C-A, Grimm D, Sugiyama M, Kawahara Y, Borgwardt KM. Efficient network-guided multi-locus association mapping with graph cuts. Bioinformatics. 2013;29: i171–i179.
8. Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein–protein interaction networks. Bioinformatics. 2011;27: 95–102.
9. Liu Y, Brossard M, Sarnowski C, Vaysse, Amaury A, Moffatt M, Margaritte-Jeannin P, et al. Network-assisted analysis of GWAS data identifies a functionally-relevant gene module for childhood-onset asthma. Sci Rep. 2017;7: 938.
10. Liu Y, Brossard M, Roqueiro D, Margaritte-Jeannin P, Sarnowski C, Bouzigon E, et al. SigMod: an exact and efficient method to identify a strongly interconnected disease-associated module in a gene network. Bioinformatics. 2017;33: 1536–1544.
11. Wang Q, Yu H, Zhao Z, Jia P. EW_dmGWAS: edge-weighted dense module search for genome-wide association studies and gene expression profiles. Bioinformatics. 2015; btv150.
12. Allum F, Hedman ÅK, Shao X, Cheung WA, Vijay J, Guénard F, et al. Dissecting features of epigenetic variants underlying cardiometabolic risk using full-resolution epigenome profiling in regulatory elements. Nat Commun. 2019;10: 1209.
13. Lea AJ, Tung J, Zhou X. A Flexible, Efficient Binomial Mixed Model for Identifying Differential DNA Methylation in Bisulfite Sequencing Data. PLoS Genet. 2015;11: e1005650.
14. Eranti P, Bouzigon E, Demenais F. Development of an analysis strategy for integrated network analysis of multi-omics data. Zenodo. 10.5281/zenodo.4321816; 2020 Dec.
15. Ferreira MA, Jansen R, Willemsen G, Penninx B, Bain LM, Vicente CT, et al. Gene-based analysis of regulatory variants identifies 4 putative novel asthma risk genes related to nucleotide synthesis and signaling. J Allergy Clin Immunol. 2017;139: 1148–1157.
16. Bakshi A, Zhu Z, Vinkhuyzen AAE, Hill WD, McRae AF, Visscher PM, et al. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. Sci Rep. 2016;6: 32894.
17. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28: 495–501.

18. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47: D607–D613.
19. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2: 100141.
20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol J R STAT SOC B. 1995; 289–300.
21. Shu L, Zhao Y, Kurt Z, Byars SG, Tukiainen T, Kettunen J, et al. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. BMC Genomics. 2016;17: 874.

Appendix

A1: Gene module associated with AD

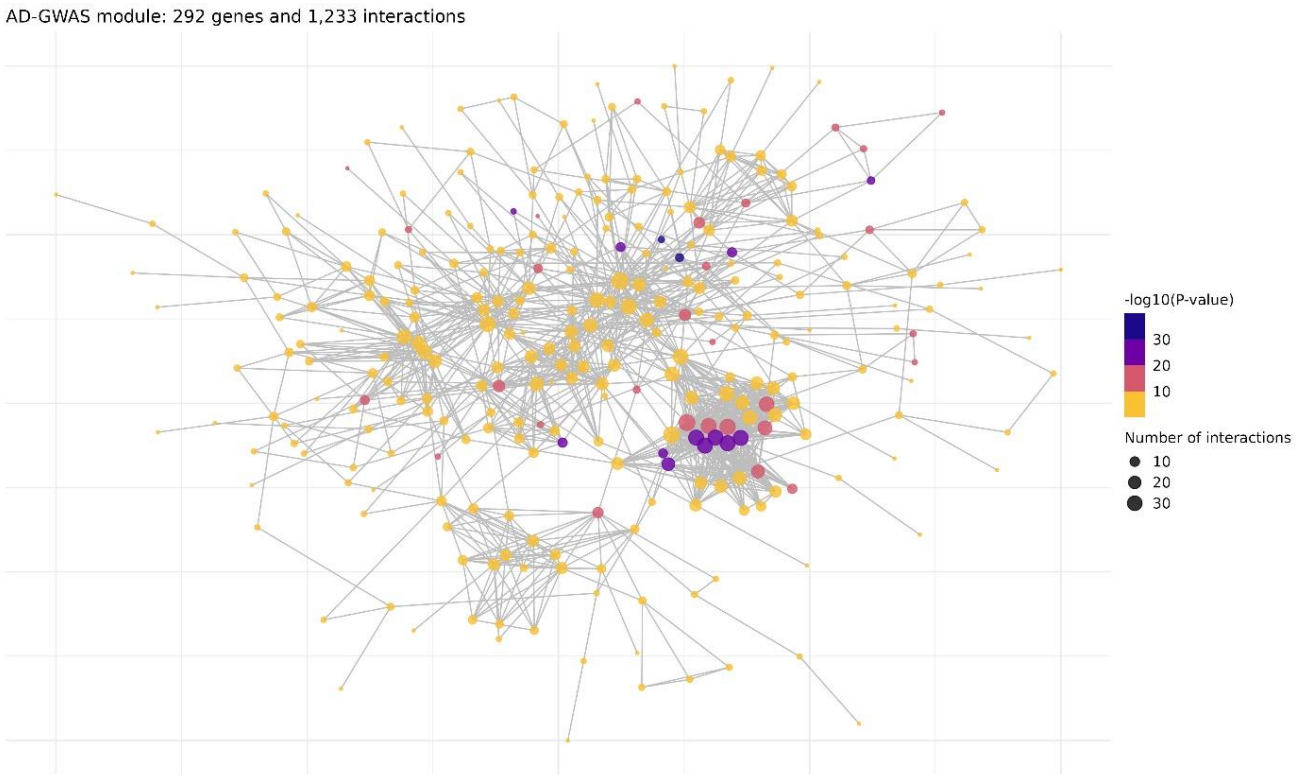

Figure S1: Pattern of interactions among the module genes identified by the network analysis of AD GWAS.

A2: Assessment of the association of AD-GWAS derived gene module with AD

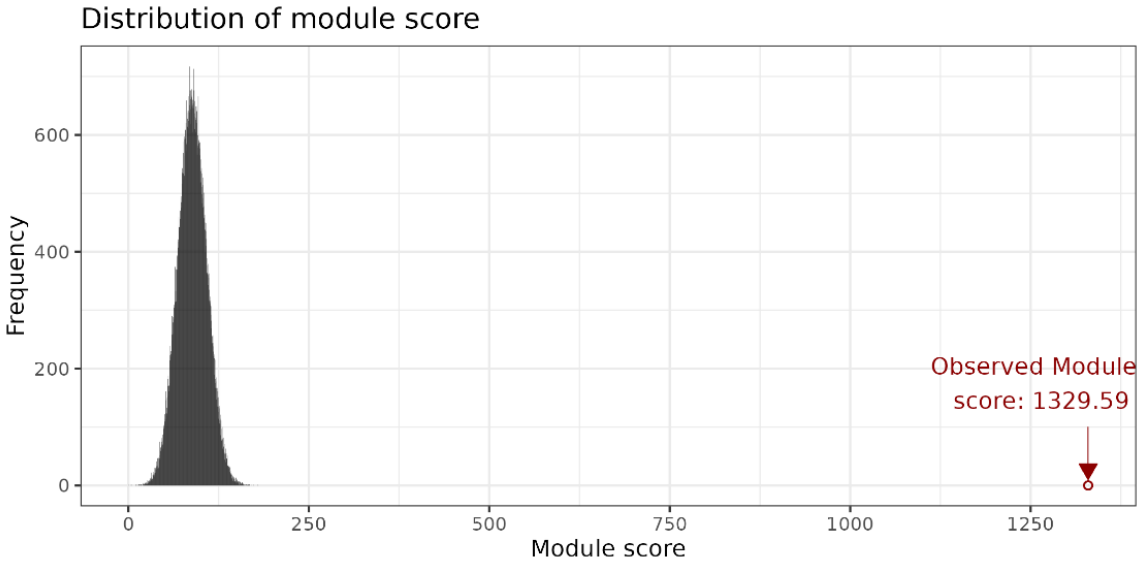

Figure S2: Histogram of the module scores of 100,000 random modules sampled from the AD-GWAS scored GeneNet. Each module has the same number of genes as the selected AD-GWAS gene module (n=292). The selected AD-GWAS gene module has a significantly higher score than the random modules (P<10-5).

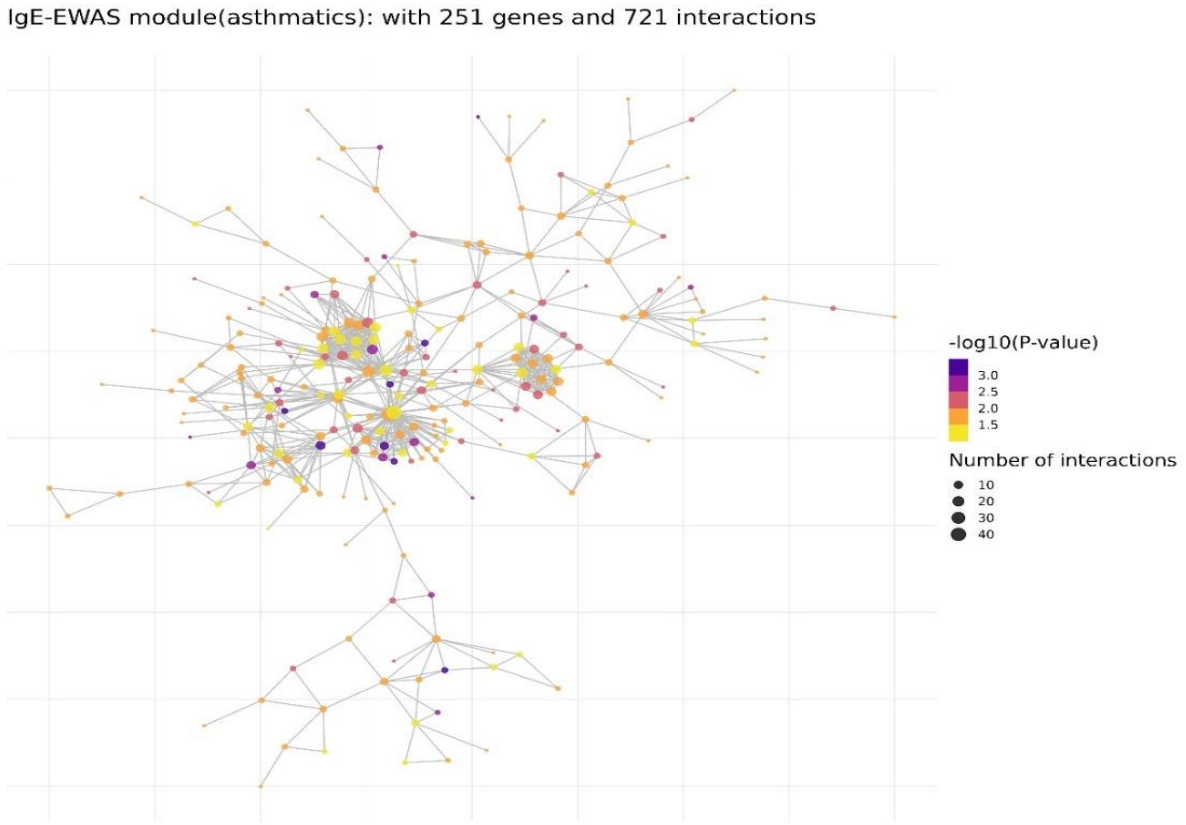

Figure S3: Pattern of interactions among the module genes identified by the network analysis of IgE-EWAS (asthmatics).

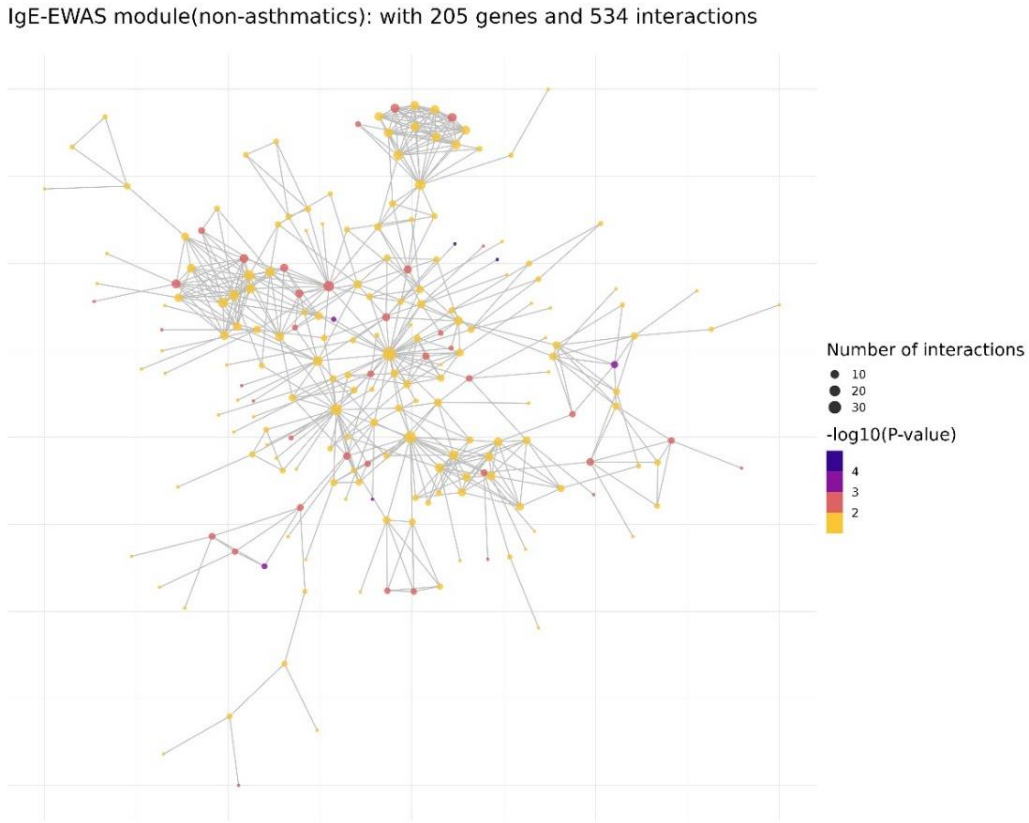

Figure S4: Pattern of interactions among the module genes identified by the network analysis of IgE EWAS (non-asthmatics).

A4: Assessment of the association of IgE-EWAS derived gene modules with IgE levels

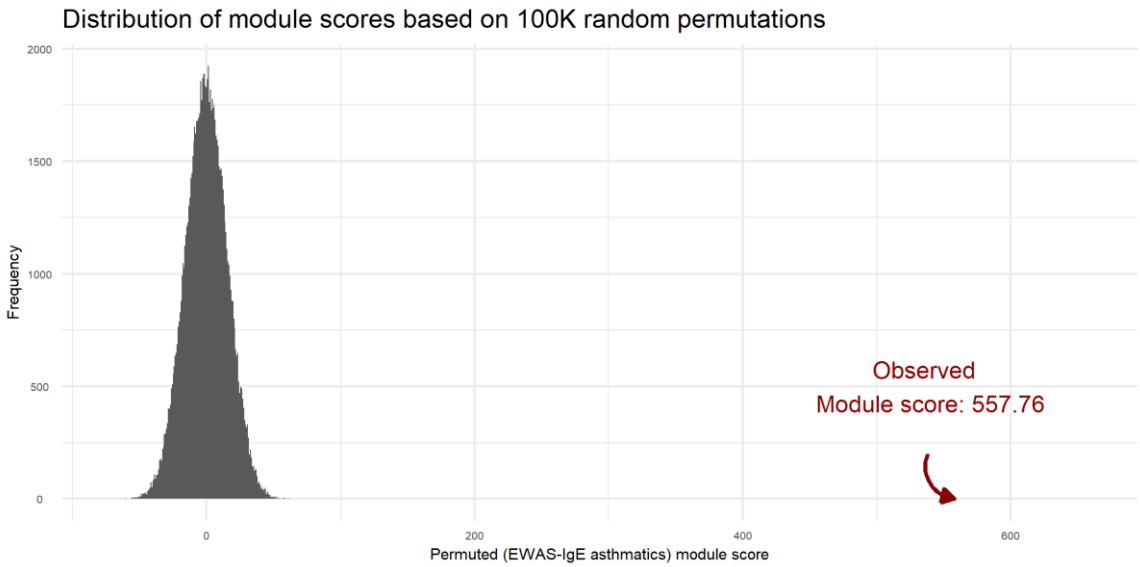

Figure S5: Histogram of the module scores of 100,000 random modules sampled from the IgE-EWAS scored GeneNet (asthmatics). Each module has the same number of genes as the selected IgE-GWAS gene module in asthmatics (n=251). The selected IgE-GWAS gene module has a significantly higher score than the random modules (P<10-5).

.

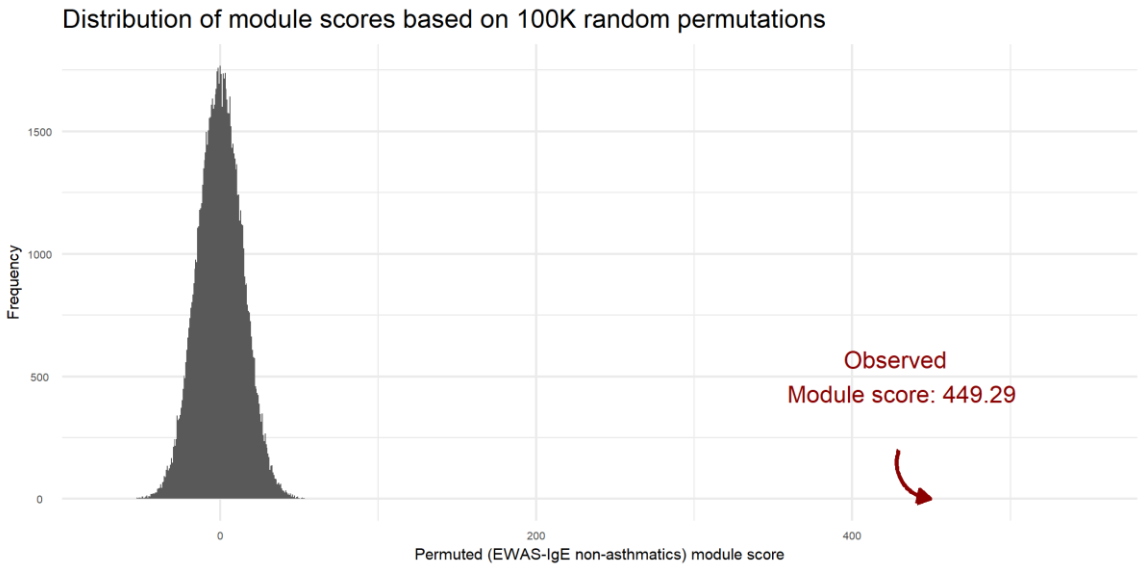

Figure S6: Histogram of the module scores of 100,000 random modules sampled from the IgE-EWAS scored GeneNet (non-asthmatics). Each module has the same number of genes as the selected IgE-GWAS gene module in non-asthmatics (n=205). The selected IgE-GWAS gene module has a significantly higher score than the random modules (P<10-5).