GENOME-WIDE GENETIC INTERACTION ANALYSIS OF GLAUCOMA USING EXPERT KNOWLEDGE DERIVED FROM HUMAN PHENOTYPE NETWORKS

TING HU†, CHRISTIAN DARABOS†, MARIA E. CRICCO, EMILY KONG, JASON H. MOORE

Institute for the Quantitative Biomedical Sciences, Geisel School of Medicine, Dartmouth College
Hanover, NH 03755, U.S.A.
E-mail: jason.h.moore@dartmouth.edu

The large volume of GWAS data poses great computational challenges for analyzing genetic interactions associated with common human diseases. We propose a computational framework for characterizing epistatic interactions among large sets of genetic attributes in GWAS data. We build the human phenotype network (HPN) and focus around a disease of interest. In this study, we use the GLAUGEN glaucoma GWAS dataset and apply the HPN as a biological knowledge-based filter to prioritize genetic variants. Then, we use the statistical epistasis network (SEN) to identify a significant connected network of pairwise epistatic interactions among the prioritized SNPs. These clearly highlight the complex genetic basis of glaucoma. Furthermore, we identify key SNPs by quantifying structural network characteristics. Through functional annotation of these key SNPs using Biofilter, a software accessing multiple publicly available human genetic data sources, we find supporting biomedical evidences linking glaucoma to an array of genetic diseases, proving our concept. We conclude by suggesting hypotheses for a better understanding of the disease.

Keywords: GWAS; Epistasis; Gene-gene interaction; Human phenotype network; Statistical epistasis network; Biofilter; Eye diseases; Glaucoma; SNPs; Pathways.

1. Introduction

With the rapid development of genotyping technologies and exponential increase in computational power, we are able to leverage the wealth of genome-wide association studies (GWAS) data to test millions of genetic variations (single nucleotide polymorphisms, SNPs) for their associations with common human diseases.[1,2] However, common disorders are usually the result of the non-linear combined effect of many variations. The study of these complex, epistatic interactions among multiple genetic attributes is crucial in explaining portions of the genotype-to-phenotype associations.[3,4]

Detecting and analyzing complex genetic interactions in GWAS data pose great statistical and computational challenges. Epistatic effects potentially involve any number of SNPs, from a couple to hundreds or even thousands. In turn, the uncertainty of size of the interacting SNP set would require interaction studies to comprehensively explore all possible combinations of SNPs. A typical GWAS dataset encompasses hundreds of thousands of variations, making the enumeration of all pairwise SNP interactions computationally infeasible, and the computational cost increases exponentially with the number of features at hand. This inevitably becomes a computational bottleneck of processing and analyzing GWAS data.

Therefore, genetic interaction studies meant to process modern GWAS data require the development of advanced informatics methodologies. These novel algorithms must be able to

†Co-first authors – TH and CD have contributed equally to this work.

simultaneously analyze large sets of interactions and identify genetic attributes potentially involved in higher-order interactions. Network modeling has emerged as a suitable framework for such purposes,[5–7] thanks to its ability to represent a large number of entities, as vertices, and their relationships, as edges.

In this study, we propose an informatics framework of detecting and analyzing genetic interactions for GWAS data. We use the example of the GLAUGEN study, a GWAS on glaucoma consisting of over a million attributes. We pre-screen the dataset using a knowledge-based filter by building human phenotype network (HPN) to prioritize SNPs and reduce the search space according to their known relationship to the disease/phenotype in question. This reduced SNP list is then used to quantitatively evaluate all pairwise epistatic interactions and to build statistical epistasis network (SEN) to characterize their global interaction structure. The key SNPs identified by the SEN are functionally annotated and evaluated for their interaction with other disorders. This new framework has the potential to elucidate large parts of the complex genetic architecture of common human diseases, such as glaucoma, and identify key SNPs for further biological validations.

2. Dataset and Methods

2.1. *Glaucoma and the GLAUGEN Study*

Glaucoma, a neurodegenerative disease, is the primary cause of irreversible blindness, affecting over 60 million people worldwide. The most common kind of glaucoma, in all populations, is primary open-angle glaucoma (POAG). Currently the only modifiable risk factor of POAG is intraocular pressure (IOP), but even lowering that will only slow the process, not stop it.[8] However, a substantial amount of POAG has been shown to have a genetic basis. Familial aggregation of POAG has been long recognized and studied to find multiple loci linked to them, causing the discovery of glaucoma-causing genes myocilin (MYOC), optineurin (OPTN) and WD[a]-repeat domain 36.[9] About 5% of POAG is presently attributed to a single-gene or Mendelian forms of glaucoma. More cases of POAG are caused by the combined epistatic effects of many genetics risk factors.[10]

The Glaucoma Gene Environment (GLAUGEN) seeks to illuminate the origin of the disease, to discover genetic loci associated with POAG, and to identify gene-gene and gene-environment associations.[8] GLAUGEN is a GWAS, case-control study, with about 2,000 unrelated cases and over 2,300 controls. To be involved in the study, both the cases and the controls had to be at least 40 years of age and European derived or Hispanic Caucasian. Subjects were genotyped with 1,048,965 SNPs examined.[8]

2.2. *Human Phenotype Network (HPN)*

The human disease network,[11] or the more general human phenotype networks (HPNs),[12,13] are mathematical graph models where nodes represent human genetic disorders and edges link those nodes with shared biology.[14] The underlying connections of the HPN contribute to

[a]tryptophan-aspartic acid

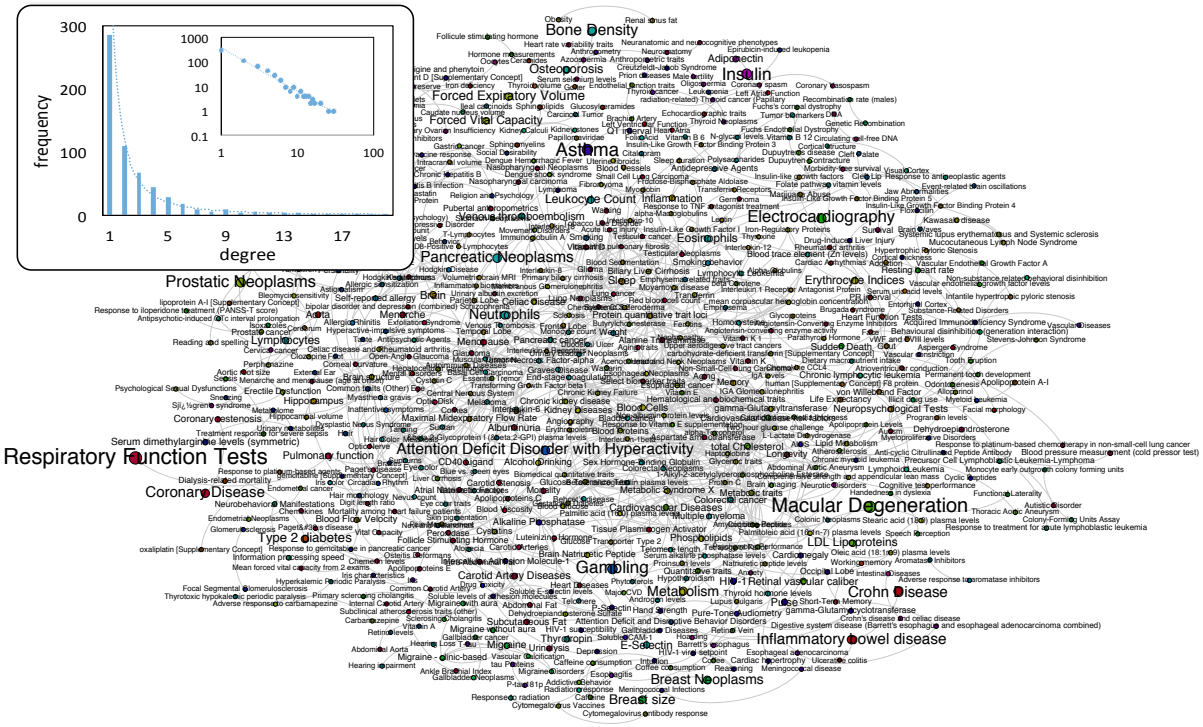


Fig. 1. The SNP-based human phenotype network, filtered (edge weight cutoff = 0.002). Vertices are colored by "modules"[13] to increase readability. The vertex sizes are proportional to the number of associated SNPs. The degree distribution is given on both linear and logarithmic scales.

the understanding of the basis of disorders, which in turn leads to a better understanding of human diseases.

In the present work, we rely on the SNP-based HPN, where connected diseases share common SNPs. We use two sources for the phenotype-to-SNPs mapping: the National Human Genome Research Institute GWAS catalog,[15] a manually curated list of all the National Institute of Heath (NIH) funded GWAS studies, and the NIH's database of Genotypes and Phenotypes (dbGaP, `http://www.ncbi.nlm.nih.gov/gap`). To build the HPN, all traits listed in the GWAS catalog and dbGaP are annotated with their risk-associated SNPs. All the traits become nodes in our network. Then, traits that have one or more SNPs in common are linked in the HPN by an edge, the weight of which is proportional to the size of the SNPs' overlap. The GWAS catalog and dbGaP report 1,252 traits combined, annotated with 37,681 SNPs in 16,411 loci. The resulting bipartite network is projected in the space of phenotype vertices to obtain the HPN shown in Fig. 1.

The HPN contains 985 vertices and over 26,000 edges, and encompasses all phenotypes listed in the GWAS catalog and dbGaP, provided that they are connected to at least one other trait. The resulting network is extremely dense, with an average degree greater than 500.

2.3. *Statistical Epistasis Network (SEN)*

Statistical epistasis network (SEN) is designed to study a global aggregation of pairwise epistatic interactions among a large number of factors.[16,17] First, pairwise epistatic interaction

is quantified using the information-theoretic measure called *information gain*[18] for all possible pairs of genetic attributes. Specifically, the genotypes of two SNPs A and B, and the discrete phenotypic class C are regarded as variables. *Mutual information* $I(A;C)$ or $I(B;C)$ measures the shared information between (A or B) and C calculated as $I(A;C) = H(A)+H(C)-H(A,C)$ or $I(B;C) = H(B)+H(C)-H(B,C)$, where *entropy* H measures the uncertainty of a random variable or joint multiple random variables. Therefore, mutual information $I(A;C)$ or $I(B;C)$ describes the reduction of the uncertainty of the phenotypic class C given the knowledge of the genotype of A or B, known as the main effect of A or B on C. Similarly, the joint mutual information $I(A,B;C)$ measures the total effect of combining A and B on explaining the phenotype C. Subtracting the individual mutual information $I(A;C)$ and $I(B;C)$ from $I(A,B;C)$ captures the gained information on C by considering A and B together rather than individually. This *information gain* $IG(A;B;C)$ is a practical measure for the epistatic interaction effect of SNPs A and B on phenotype C, and has been used as an efficient non-parametric and model-free statistical quantification of pairwise epistasis in genetic association studies.[19–23]

Second, all pairwise SNP-SNP interactions are quantified and ranked. We build statistical epistasis networks where vertices are SNPs and edges are weighted pairwise interactions. We only include pairs of SNPs if their interaction strengths are stronger than a preset threshold. We analyze the network topological properties at each cutoff value of the threshold, such as the size of the network (the number of its vertices and the number of its edges), the connectivity of the network (the size of its largest connected component), and its vertex degree distribution.

Then, a threshold of the pairwise interaction strength is determined systematically by finding the cutoff where the topological properties of the real-data network differentiate the most from the null distribution of permutation testing. Such a SEN provides a significant global structure of clustered strong pairwise epistatic interactions associated with a particular phenotype. It serves as a map for further network properties investigation and key SNPs prioritization as discussed in the following section.

2.4. *Network Property Analysis of SEN*

The *assortativity* of a network measures the propensities of vertices with similar characteristics to connect to one another.[24,25] In the context of SENs, we are interested in looking into the main effect assortativity, i.e., whether there exists a correlation of main effects between pairs of interacting SNPs. Such main effect assortativity is calculated as the Pearson correlation coefficient r of the main effects at either ends of an edge in a SEN,

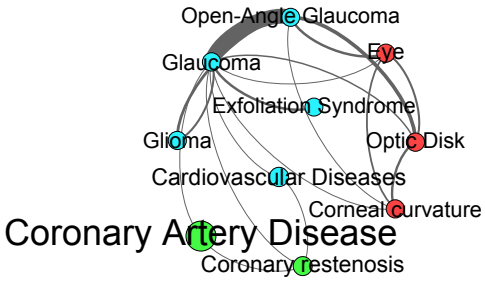
$$r = \frac{M^{-1}\sum_i j_i k_i - [M^{-1}\sum_i \frac{j_i+k_i}{2}]^2}{M^{-1}\sum_i \frac{j_i^2+k_i^2}{2} - [M^{-1}\sum_i \frac{j_i+k_i}{2}]^2}, \quad (1)$$

where M is the total number of edges, and j_i and k_i are the main effects, calculated as the mutual information of a SNP and the phenotypic class of the vertices (SNPs) at the ends of the i-th edge, with $i = 1, 2, ..., M$. The coefficient r lies between -1 and 1, with $r =$ 1 indicating perfectly assortative, $r = 0$ for non-assortative, and $r = -1$ for complete disassortative networks.

At the vertex level, *centrality* measures the importance of individual vertices in a network. The most commonly used centrality measure is the node's *degree* $C_D(v)$, which is the total

Phenotype	#SNPs	Degree
Exfoliation Syndrome	1	1
Coronary Artery Disease	639	2
Cardiovascular Diseases	70	2
Corneal curvature	13	4
Optic Disk	17	4
Open-Angle Glaucoma	6	5
Glioma	12	2
Eye	13	4
Glaucoma	18	9
Coronary restenosis	56	3

(a)



(b)

Fig. 2. Glaucoma network neighborhood. (a) Phenotypes, including the number of SNPs associated with each phenotype and the "strength" of the interaction (degree) with the rest of the sub-network. (b) a depicted representation of the Glaucoma-centered sub-network, in which the node sizes are proportional to the number of associated SNPs.

number of edges connected to vertex v. The degree of a SNP in the statistical epistasis network shows the number of other SNPs with which it is interacting. *Betweenness* centrality is a more sophisticated metric that quantifies the number of times a vertex is part of the shortest path between any pair of vertices,[26] represented as $C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where σ_{st} is the total number of shortest paths from vertex s to vertex t and $\sigma_{st}(v)$ is the number of those paths that pass through vertex v. *Closeness* centrality, defined as $C_C(v) = \frac{1}{\sum_{s \neq v} d_{vs}}$, where d_{vs} is the distance between vertices v and s.[27,28] This metric describes how easily a given vertex can reach all other vertices in a network. In the context of SENs, centrality measures are used to identify key SNPs that play an essential role in the global interaction structure.

3. Results

3.1. *Pre-screening of Glaucoma Data Using HPN*

To reduce the computational cost (in time and resources) of analyzing a GWAS dataset, the data must be filtered, selected and/or prioritized prior to processing. Algorithmic filtering methods fall into two main families: knowledge-driven and data-driven. Knowledge-driven filters relies on the actual known biology behind the data, whereas data-driven filtering solely relies on the statistical pre-processing of the data, regardless of their type. ReliefF, SURF, and TuRF[29,30] are examples of data-driven statistical methods of preprocessing GWAS data. For a complete review of the knowledge-based vs. data-driven filtering, refer to Sun *et al.*[31]

The method we propose to use here starts with the full HPN described in Section 2.2. We identify the glaucoma vertex and establish a list of the disorders and phenotypes in its immediate neighborhood. The 10 members of the "glaucoma neighborhood" are presented in table in Fig. 2.

Using the sub-HPN, we compile a list of 948 SNPs that have been associated with any of the phenotypes in the sub-network. We subsequently annotate each of the risk-associated SNPs with its linkage disequilibrium[32] SNPs, indirectly associated with an increased risk. The full list of 4,388 features containing both direct and indirect risk-associated SNPs is used to filter the full one million SNPs of the GLAUGEN GWAS data. The subset made of the overlap between our priority-list and the GLAUGEN dataset, composed of 2,380 SNPs, is

used to build the SNP interaction network presented in the following section.

3.2. *Statistical Epistasis Network of Glaucoma*

Using the list of 2,380 SNPs, we evaluate all the pairwise epistatic interactions (>2.8 million) using the information gain measure. The strongest interacting pair is SNPs rs13123790 and rs6572380 with a strength of 1.163% association of the disease class. We set the pairwise interaction strength threshold decreasing from the strongest observed value with a step of 0.0001. At each cutoff, a SEN is built by including pairs of SNPs as edges and two end vertices if their interaction strength is above the cutoff. Then for each network we evaluate its topological properties including the number of non-isolated vertices, the number of edges, the sum of weighted edges, and the size of its largest connected component. A 100-fold permutation test assesses the significance of these observed network properties. We permute the data 100 times by randomly shuffling the disease class column, and for each permuted dataset all the pairwise interaction strengths are calculated and networks are built using the same cutoffs as the read-data networks.

We observe a network at cutoff 0.693% that has a largest connected component including significantly more vertices ($p = 0.05$) than those permuted-data networks at the same cutoff. This largest connected component has 713 vertices and 789 edges (Fig. 3).

The main effect assortativity of this network is -0.053 with a significance of $p = 0.04$ using a 100-fold edge swapping permutation test where we randomly pick $10 \times |E|$ ($|E|$ is the total number of edges) pairs of edges and swap their end vertices. This significant negative assortativity indicates that the main effects of interacting SNPs are negatively correlated, i.e., high main-effect vertices tend to interact more with low main-effect vertices.

The degree, betweenness, and closeness centralities of all the vertices in the network are evaluated, and their distributions are shown in Fig. 4. Most of vertices have degrees equal to or less than 3. However there are a minority vertices with a degree higher than 5, meaning that they interact with larger number of other SNPs. In the distribution of betweenness centrality, the majority of vertices have low betweenness. However, some vertices have a much higher betweenness than the rest. The closeness centrality follows a normal distribution indicating that most vertices have similar access to all other vertices. We sort the vertices by descending centrality. At the top, we find the key SNPs that are potentially involved the genetic processes responsible for glaucoma. We further investigate the clinical and biomedical implications of these SNPs through Biofilter[33] functional annotations. Biofilter provides a single interface to multiple publicly available online human genetic datasources (`https://ritchielab.psu.edu/software/biofilter-download`).

4. Discussion

In this study, we proposed a genetic interaction analysis framework of using human phenotype network (HPN) as knowledge-based data filter and subsequently statistical epistasis network (SEN) for quantitative epistatic interaction detection and visualization. Our method was successfully applied to GLAUGEN GWAS data and we were able to identify a connected network of interacting SNPs that included significantly more SNPs than expected randomly. Such a

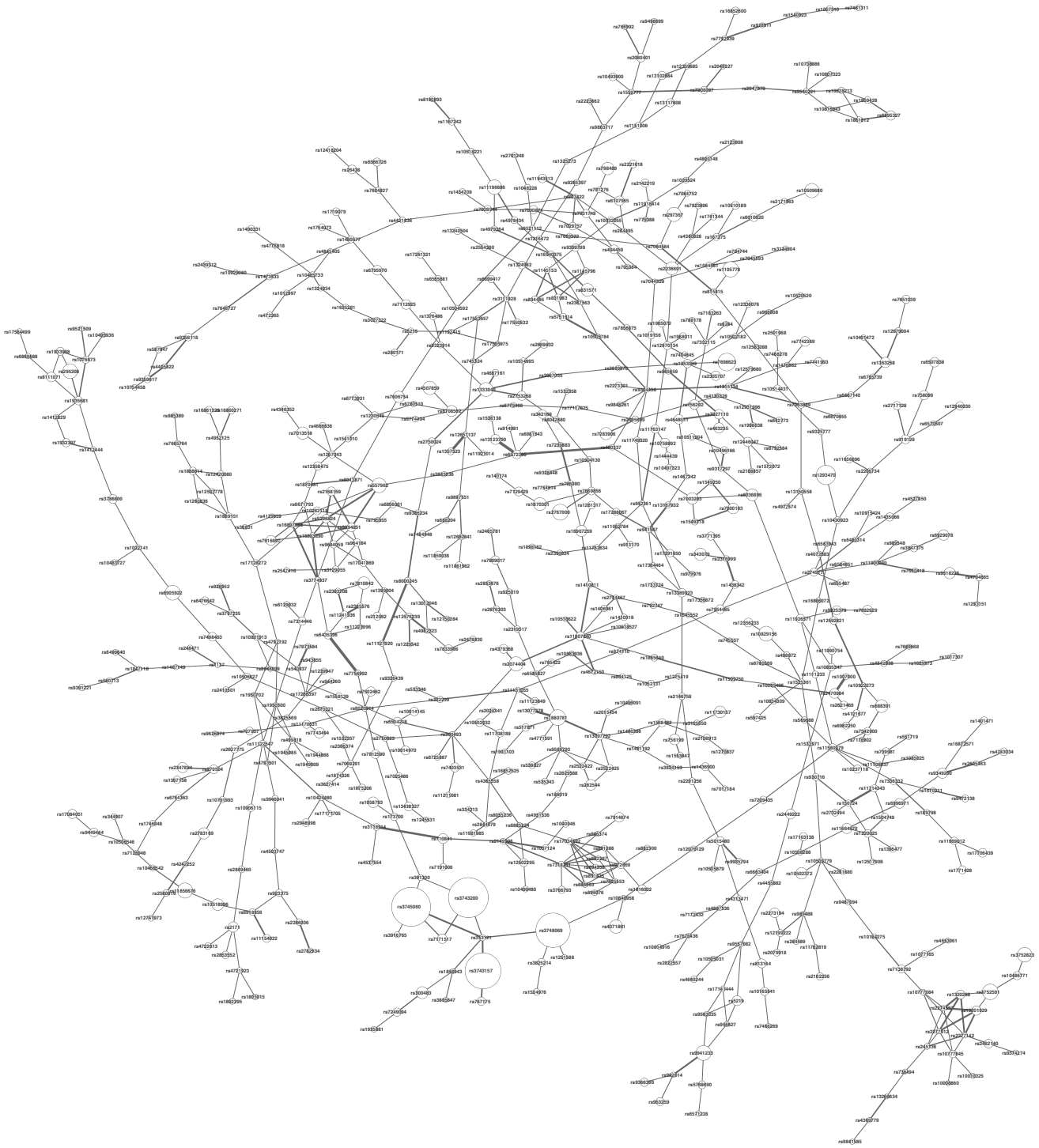


Fig. 3. The SEN of Glaucoma. The network includes 713 SNPs (vertices) and 789 pairwise interactions (edges). The size of a vertex represents the strength of the main effect of its corresponding SNP, with the disease association ranging from $\ll$0.001% to 1.124%. The width of an edge indicates the strength of its corresponding interaction, with the disease association ranging from 0.693% to 1.163%.

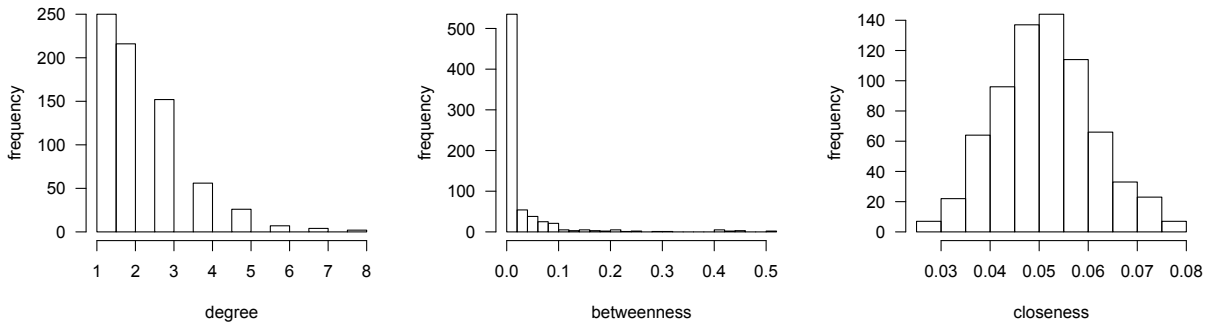


Fig. 4. The distributions of (a) degree, (b) betweenness, and (c) closeness centralities of vertices in the SEN of Glaucoma.

large connected SNP network indicated a complex genetic basis of glaucoma.

HPN is an computational model that systematically organizes and clusters various phenotypes and diseases based on their shared genetic association factors or related basic functional units. It also serves a powerful knowledge-based filter for GWAS data pre-screening and prioritization. Indeed, when compared to established data-driven filters, including the family of Relief algorithms, the HPN filtering performed equally well. This was evaluated by quantifying all the pairwise interactions of filtered sets of SNPs using different data-driven filters (data not shown).

SEN is able to detect and analyze genetic interactions among a large set of attributes in genetic association studies. The derived SNP interaction network of glaucoma in this study had a significant connected structure and may serve as a map for identifying key genetic factors associated with the disease with further biological validations. The significant negative main effect assortativity of such a network indicated that epistatic interactions were likely to happen between SNPs with negatively correlated main effects. This finding suggests that most existing main-effect-based GWAS data filtering algorithms may overlook potential important interactions since some SNPs with low main effects would not be pre-selected using such filters.

The set of SNPs with highest ranked centralities were annotated using Biofilter[33] to their associated genes and, in turn, the genes to pathways and functional groups. We used the annotation to identify other complex diseases possibly sharing biology with glaucoma and suggest novel hypotheses for undocumented shared genetic interactions.

It is known that glaucoma and type II diabetes (T2D) are associated, but direct genetic relationship remains unknown. We found SNPs connecting T2D to glaucoma: rs1333040 ($C_D = 5, C_B = 0.043$) and rs1053049 ($C_D = 5$). The genes CDKN2B-AS1 and PPARD, which these two SNPs are respectively located in, are known to be associated with diabetes. Peng *et al.* observed that CDKN2A/B gene had a 1.17-fold (95% CI: 1.1-1.23) risk estimate for T2D.[34] PPARD has been found to be associated with higher fasting glucose, glycated hemoglobin (HbA1c), and increased risk of T2D and combined impaired fasting glucose/T2D in a population-based Chinese Han sample.[35] SNP rs9540221 ($C_D = 5$) is also intergenic between NFYAP1 and

LGMNP1, of which the former has also been associated with diabetes and obesity. Indeed, population-based studies suggest that persons with T2D have an increased risk of developing open-angle glaucoma (OAG): the Beaver Dam Eye Study,[36] the Framingham Eye Study,[37] and the Los Angeles Latino Eye Study.[38] Chopra *et al.* has noted that there is a positive association between T2D and OAG in a sample Latino population study in Los Angeles, USA.[38] The Rotterdam Study[39] has also previously reported a positive association between DM and OAG, with a relative risk of 3.11 (95% CI, 1.12-8.66). Laboratory evidence for T2D and glaucoma linkages also exists. Some have observed that there is an inner retinal cell death, determined by the presence of hyaline bodies in the ganglion cell and retinal nerve fiber layer of the eyes of post-mortem eyes of patients with T2D.[40] Possible hypotheses of T2D to glaucoma direct linkage include how altered biochemical pathways and vascular changes in T2D reduce blood flow and impaired oxygen diffusion, thus increasing oxidative stress and endothelial cell injury.[41] Excessive glial cell activation may contribute to chronic inflammation, making the retinal ganglion cells more susceptible to glaucomatous damage.[42] Connective tissue remodeling might also promote increased intraocular pressure (IOP) and greater optic nerve head mechanical stress, which can result in glaucomatous optic neuropathy.[43] Diabetes is known to cause microvascular damage and may affect the vascular autoregulation of the optic nerve and retina. Thus a longer duration of T2D would be associated with a higher risk of OAG.

Obesity, like T2D, is a trait often associated with glaucoma, but whose connection remains elusive. We found associations between obesity and glaucoma in SNPs rs12970134 ($C_D = 5, C_C = 0.062$), rs11807640 ($C_D = 8, C_B = 0.516, C_C = 0.071$), rs9540221 ($C_D = 5$), rs10777845 ($C_D = 5$), and rs4365558 ($C_D = 5, C_C = 0.063$), all of which showed significance in our glaucoma study. SNP rs12970134 is located upstream of gene MC4R, the defects of which have been identified as a cause for autosomal dominant obesity.[44] SNP rs11807640 is located downstream to gene CDK4PS, which has been associated with obesity traits among some postmenuposal women.[45] Both rs9540221, which is intergenic between NFYAP1 and LGMNP1 and rs4365558, which is intergenic between A4GALT and RPL5P34, are SNPs situated between genes that have been associated with obesity as well. Finally, rs10777845 is upstream to gene RMST, which has been observed to be associated with severe early-onset obesity.[46] In the investigation to determine direct linkage between glaucoma and obesity, several hypotheses exist. Newman-Casey *et al.* observed that not only was the frequency of OAG was higher among obese individuals (3.1%) than among non-obese individuals (2.5%), but that the relationship between OAG and obesity had significant correlations with gender.[47] Obese women had a 6% increased hazard of developing OAG when compared to non-obese women, while there was no significant effect in men. Similar findings were made by Zang and Wynder, in which they observed OAG diagnosis as twice as likely in women with a body mass index (BMI) above 27.5, but no impact of BMI in men.[48] Some hypothesize that increased orbital pressure as a result of excess fat tissue may cause a rise in venous pressure and a consequent increase in intraocular pressure (IOP).[47]

Our study also show links between Alzheimer's disease (AD) and glaucoma. The SNP rs803422 ($C_D = 5, C_C = 0.060$), which we found to affect glaucoma is also associated with

AD. In addition, the two genes SORCS3 and MTHFD1L which contain two of the SNPs with the strongest association are also known to be associated with AD. Patients with AD have a significantly increased rate of glaucoma occurrence. In a study in four nursing homes in Germany, 112 Alzheimer's patients were taken as a case group. 29 of those 112 were found to have Glaucoma, a rate of 25.9% as opposed to a 5.2% rate in the control group.[49] However, in spite of significant research on this topic, there have still been no clear clinical or genetic relationships found between the two diseases.[50]

Perhaps unsurprisingly, our research has also led to us connecting glaucoma with other phenotypes relating to the eye, particularly cataracts and myopia. The association between cataracts and glaucoma can be seen in the similar genes CDK4PS , NFYAP1, and LGMNP1, all of which contain SNPs found to be highly associated with glaucoma. Myopia and glaucoma are both directly related to the SNP rs569688 ($C_B = 0.415, C_C = 0.074$). There have been many recent studies about the connection between glaucoma and myopia, particularly primary open-angle glaucoma (POAG) and high myopia. Though the connection has still not been found, it is believed the high myopia is another risk factor of POAG, though it is not as important as intraocular pressure. However, it is recommended that those with high myopia be screen for glaucoma more frequently.[51]

Of the SNPs and genes associated to glaucoma in our study, the majority that are not directly eye-related, are heart related. Glaucoma is connected therefore probably connected to cardiovascular disease, coronary artery disease, and coronary restenosis. SNPs rs17034592 ($C_D = 7, C_C = 0.060$) and loci LDB2, CDKN2B and TRNAQ46P are all associated with coronary artery disease. SNPs rs499818 ($C_D = 5, C_C = 0.052$), rs1333040 ($C_D = 5, C_C = 0.052$), and loci A4GALT, RPL5P34, and CST7 are all associated with myocardial infarction, and A4GALT and RPL5P34 are associated with cardiovascular disease. Studies have shown that cardiovascular disease may play an important role in the development and progression of glaucoma. However, it is also hypothesized that this relationship may be indirect, due to the variable of age.[52] Additional studies about the relationship between intraocular pressure, the main risk factor for glaucoma, and cardiovascular disease exist. However, any genetic relationship between these two diseases is yet to be found.

An undocumented interaction found in this study is the link between glaucoma and colorectal carcinoma. Relevant SNPs that were significant were rs972869 ($C_D = 7, C_C = 0.052$), located near the EPHB1 gene, and rs300493 ($C_D = 6, C_B = 0.311, C_C = 0.066$), located in the gene NAV3, both of which are linked to colorectal carcinoma. Studies have shown that obesity and T2D are associated risks of colorectal cancer, which could lead to the hypothesis that colorectal carcinoma and glaucoma may be co-morbidities and are affected by similar pathways.

In summary, the successful application of our methodology on GLAUGEN GWAS data proves the effectiveness of our bioinformatics framework, and suggests its great potential in detecting and characterizing gene-gene interactions in GWAS data. In future studies, we expect to extend the interaction analysis beyond the order of pairs. The increased computational demand would, however, require a stricter data pre-filtering. The more stringent SNP filtering can be achieved using a HPN based on SNP overlap to restrict the interactions to the most

significant ones. Other avenues of research would include using eye-specific tissue genotype data to further refine the analysis and the drug response of the nodes of interest.

Acknowledgements

This work was supported by the National Institute of Health (NIH) grants R01-EY022300, R01-LM010098, R01-LM009012, R01-AI59694, P20-GM103506, P20-GM103534.

References

1. R. Sachidanandam, D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein and et al., *Nature* **409**, 928 (2001).
2. W. Y. S. Wang, B. J. Barratt, D. G. Clayton and J. A. Todd, *Nature Review Genetics* **6**, 109 (2005).
3. H. J. Cordell, *Human Molecular Genetics* **11**, 2463 (2002).
4. J. H. Moore, *Nature Genetics* **37**, 13 (2005).
5. N. A. Davis, J. E. Crowe Jr., N. M. Pajewski and B. A. McKinney, *Genes and Immunity* **11**, 630 (2010).
6. T. Hu and J. H. Moore, Network modeling of statistical epistasis, in *Biological Knowledge Discovery Handbook: Preprocessing, Mining, and Postprocessing of Biological Data*, eds. M. Elloumi and A. Y. Zomaya (Wiley, 2013) pp. 175–190.
7. A. Pandey, N. A. Davis, B. C. White, N. M. Pajewski, J. Savitz, W. C. Drevets and B. A. McKinney, *Translational Psychiatry* **2**, p. e154 (2012).
8. J. L. Wiggs, M. A. Hauser, W. Abdrabou, R. R. Allingham, D. L. Budenz, E. Delbono, D. S. Friedman, J. H. Kang, D. Gaasterland, T. Gaasterland, R. K. Lee, P. R. Lichter, S. Loomis, Y. Liu, C. McCarty, F. A. Medeiros, S. E. Moroi, L. M. Olson, A. Realini, J. E. Richards, F. W. Rozsa, J. S. Schuman, K. Singh, J. D. Stein, D. Vollrath, R. N. Weinreb, G. Wollstein, B. L. Yaspan, S. Yoneyama, D. Zack, K. Zhang, M. Pericak-Vance, L. R. Pasquale and J. L. Haines, *J Glaucoma* **22**, 517 (Sep 2013).
9. M. Takamoto and M. Araie, *Ophthalmology Journal* **58**, 1 (2014).
10. J. H. Fingert, *Eye (Lond)* **25**, 587 (May 2011).
11. K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal and A.-L. Barabasi, *Proceedings of the National Academy of Sciences* **104**, 8685 (2007).
12. C. Darabos, K. Desai, R. Cowper-Sallari, M. Giacobini, B. Graham, M. Lupien and J. Moore, *Lecture Notes in Computer Science* **7833**, 23 (2013).
13. C. Darabos, M. J. White, B. E. Graham, D. N. Leung, S. Williams and J. H. Moore, *BioData Mining* **7**, p. 1 (Jan 2014).
14. H. Li, Y. Lee, J. L. Chen, E. Rebman, J. Li and Y. A. Lussier, *Journal of the American Medical Informatics Association : JAMIA* **19**, 295 (January 2012).
15. D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff and H. Parkinson, *Nucleic Acids Res* **42**, D1001 (Jan 2014).
16. T. Hu, N. A. Sinnott-Armstrong, J. W. Kiralis, A. S. Andrew, M. R. Karagas and J. H. Moore, *BMC Bioinformatics* **12**, p. 364 (2011).
17. T. Hu, Y. Chen, J. W. Kiralis and J. H. Moore, *Genetic Epidemiology* **37**, 283 (2013).
18. T. M. Cover and J. A. Thomas, *Elements of Information Theory: Second Edition* (Wiley, 2006).
19. A. Jakulin and I. Bratko, Analyzing attribute dependencies, in *Proceedings of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2003)*, , Lecture Notes in Artificial Intelligence Vol. 2838 (Springer-Verlag, 2003).

20. J. H. Moore, J. C. Gilbert, C.-T. Tsai, F.-T. Chiang, T. Holden, N. Barney and B. C. White, *Journal of Theoretical Biology* **241**, 252 (2006).
21. D. Anastassiou, *Molecular Systems Biology* **3**, p. 83 (2007).
22. J. H. Moore, N. Barney, C.-T. Tsai, F.-T. Chiang, J. Gui and B. C. White, *Human Heredity* **63**, 120 (2007).
23. B. A. McKinney, J. E. Crowe, J. Guo and D. Tian, *PLoS Genetics* **5**, p. e1000432 (2009).
24. M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, 2010).
25. M. E. J. Newman, *Physical Review Letters* **89**, p. 208701 (2002).
26. L. C. Freeman, *Sociometry* **40**, 35 (1977).
27. A. Bavelas, *Journal of the Acoustical Society of America* **22**, 725 (1950).
28. G. Sabidussi, *Psychometrika* **31**, 581 (1966).
29. J. H. Moore and B. C. White, *Lecture Notes in Computer Science* **4447**, 166 (2007).
30. C. S. Greene, N. M. Penrod, J. Kiralis and J. H. Moore, *BioData Min* **2**, p. 5 (2009).
31. X. Sun, Q. Lu, S. Mukheerjee, P. K. Crane, R. Elston and M. D. Ritchie, *Front Genet* **5**, p. 106 (2014).
32. C. International HapMap, *Nature* **437**, 1299 (Oct 2005).
33. S. A. Pendergrass, A. Frase, J. Wallace, D. Wolfe, N. Katiyar, C. Moore and M. D. Ritchie, *BioData Min* **6**, p. 25 (2013).
34. F. Peng, D. Hu, C. Gu, X. Li, Y. Li, N. Jia, S. Chu, J. Lin and W. Niu, *Gene* **531**, 435 (Dec 2013).
35. L. Lu, Y. Wu, Q. Qi, C. Liu, W. Gan, J. Zhu, H. Li and X. Lin, *PLoS One* **7**, p. e34895 (2012).
36. B. E. Klein, R. Klein and S. C. Jensen, *Ophthalmology* **101**, 1173 (Jul 1994).
37. P. Mitchell, W. Smith, T. Chey and P. R. Healey, *Ophthalmology* **104**, 712 (Apr 1997).
38. V. Chopra, R. Varma, B. A. Francis, J. Wu, M. Torres and S. P. Azen, *Ophthalmology* **115**, 227 (Feb 2008).
39. S. de Voogd, M. K. Ikram, R. C. W. Wolfs, N. M. Jansonius, J. C. M. Witteman, A. Hofman and P. T. V. M. de Jong, *Ophthalmology* **113**, 1827 (Oct 2006).
40. J. R. Wolter, *Am J Ophthalmol* **51**, 1123 (May 1961).
41. V. H. Y. Wong, B. V. Bui and A. J. Vingrys, *Clin Exp Optom* **94**, 4 (Jan 2011).
42. M. Nakamura, A. Kanamori and A. Negi, *Ophthalmologica* **219**, 1 (Jan-Feb 2005).
43. V. C. Lima, T. S. Prata, C. G. V. De Moraes, J. Kim, W. Seiple, R. B. Rosen, J. M. Liebmann and R. Ritch, *Br J Ophthalmol* **94**, 64 (Jan 2010).
44. K. M. Ling Tan, S. Q. D. Ooi, S. G. Ong, C. S. Kwan, R. M. E. Chan, L. K. Seng Poh, J. Mendoza, C. K. Heng, K. Y. Loke and Y. S. Lee, *J Clin Endocrinol Metab* **99**, E931 (May 2014).
45. P. Coveney and R. Highfield, *Frontiers of Complexity: The Search for Order in a Chaotic World* (Faber and Faber, London, 1995).
46. E. Wheeler, N. Huang, E. G. Bochukova, J. M. Keogh, S. Lindsay, S. Garg, E. Henning, H. Blackburn, R. J. F. Loos, N. J. Wareham, S. O'Rahilly, M. E. Hurles, I. Barroso and I. S. Farooqi, *Nat Genet* **45**, 513 (May 2013).
47. P. A. Newman-Casey, N. Talwar, B. Nan, D. C. Musch and J. D. Stein, *Ophthalmology* **118**, 1318 (Jul 2011).
48. E. A. Zang and E. L. Wynder, *Nutr Cancer* **21**, 247 (1994).
49. A. U. Bayer, F. Ferrari and C. Erb, *Eur Neurol* **47**, 165 (2002).
50. P. Wostyn, K. Audenaert and P. P. De Deyn, *Br J Ophthalmol* **93**, 1557 (Dec 2009).
51. S.-J. Chen, P. Lu, W.-F. Zhang and J.-H. Lu, *Int J Ophthalmol* **5**, 750 (2012).
52. S. S. Hayreh, *Survey of Ophthalmology* **43, Supplement 1**, S27 (1999).