Computers in Biology and Medicine 109 (2019) 311–321

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Systematic analysis of genes and diseases using PheWAS-Associated networks

Ali Khosravi[a], Morteza Kouhsar[a], Bahram Goliaei[b], B. Jayaram[c], Ali Masoudi-Nejad[a,*]

[a] Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran
[b] Department of Biophysics & Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran
[c] Department of Chemistry & Supercomputing Facility for Bioinformatics & Computational Biology, Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, India

ARTICLE INFO

Keywords:
PheWAS associations
Associated gene network
Associated disease network
Systems biology

ABSTRACT

Several scientific sources have reported different causes of various diseases. One of these factors is genetic variation. Natural selection, molecular evolution and susceptibility to external conditions are the main causes of genetic variations. Phenome-Wide Association Studies (PheWAS) can emphasize the associations of genetic variations and diseases. The systematic analysis of these associations can highlight various important aspects of gene correlations and disease relationships. In this study, we have investigated a systematic approach to analyze associated networks of genes and diseases to explore novel scientific information. We have constructed the Associated Gene Network (AGN, n = 1769) and the Associated Disease Network (ADN, n = 503) based on common diseases and genes, respectively. We have evaluated these networks based on topological measures and compared them with a randomized null network. The comparative modular analysis based on size and quantity is a clear indication of the significance of these networks. We have found numerous novel associations of genes involved in different diseases. We have also found different diseases related to one another, which can correlate scientific evidence. We have verified our analysis through GO and KEGG enrichment for different case studies and concluded that AGN and ADN can be used as reference biological networks for various purposes such as drug design and drug repurposing.

1. Introduction

Enormous progress has been made in the identification of complex trait susceptibility loci in the past few years. High-throughput sequencing and RNA-Seq data have dramatically improved and the cost of analysis has reduced intensively. In this regard, the identification of genetic variations has improved Genome-Wide Association Studies (GWAS), which can identify DNA variants and phenotypic trait associations. As phenotypic traits can be caused by various variants/SNPs, GWAS can identify influencing alleles and genes for all diseases. In contrast, analysis of any single allele from significant SNPs in the form of pleiotropy can be associated with one or a group of phenotypic traits through Phenome-Wide Association Studies (PheWAS) [1]. Unlike GWAS, PheWAS can determine the possibility of various diseases using a single allele from a specific gene. As the given approach can identify several novel disease-gene associations, many scientific communities have been attracted to this method. Hence, the entire phenotypic trait that can be manipulated by SNP mutants could be identified.

Denny and his group [2] analyzed genetic variants from genome-wide association data with longitudinal electronic medical records (EMRs) to provide disease–gene associations, which was called a phenome-wide association study [3]. PheWAS has promoted a new type of association between genes, which are related to analyzed SNPs, and phenotypic traits. The application of PheWAS is generating growing sets of variant data from genomic markers associated with an increased risk for a multitude of different diseases [4]. PheWAS has provided the detailed analysis of physiopathology with the help of these variations. This systematic analysis can identify various associations of genetic variation and disease etiologies [5,6].

In this study, PheWAS data has been retrieved from the PheWAS catalog, a systematic catalog, that originates from electronic medical record systems to correlate genetic variants derived from massive cohorts and classified by International Classification of Diseases version 9 (ICD9) codes. ICD9 is used to describe discrete phenotypes [7]. Gene-disease association by GWAS and PheWAS can influence novel relationships between gene-targets and can be effective in drug

* Corresponding author.
E-mail address: amasoudin@ut.ac.ir (A. Masoudi-Nejad).

https://doi.org/10.1016/j.compbiomed.2019.04.037
Received 4 March 2019; Received in revised form 28 April 2019; Accepted 28 April 2019
0010-4825/

development [8]. The potential impact of GWAS in the prediction of drug repositioning [9–12] can also be observed through PheWAS [1,13,14].

In this study, we have constructed human association networks using revealed associations based on the PheWAS catalog. We have reconstructed two complementary networks, the Associated Gene Network (AGN) and Associated Disease Network (ADN), from the associations of different genes from common diseases and diseases with common genes, respectively. Furthermore, we have shown the significance of the above networks based on statistical and modular analysis. We have also studied the biological implications of the AGN and ADN networks with the help of a few case studies.

We also enriched the AGN and its subnetworks based on Gene Ontology. We observed the consequences of GO enrichment and novel genetic associations of similar biological processes. We have also identified relationships among diseases based on the modular analysis of the ADN. These analyses unequivocally establish the utility of PheWAS for examining Disease-Gene associations and their roles in the human disease network.

2. Methods

2.1. Data description

We collected a list of disease-associated genes in humans that are organized in a PheWAS catalog and published in Nature Biotechnology [15]. The complex disease-gene catalog reports genetic variation by p-value less than or equal to 0.05 as a conventional threshold value in GWAS and PheWAS studies [16–18]. The dataset contained more than 215K SNPs and more than 1700 unique genes associated with more than 1350 ICD9 diseases. We verified the disease classes in ICD9 by matching them with MeSH (Medical Subject Headings) terms to standardize our disease nomenclature [19,20] and excluded the list of unmatched diseases. We have removed all unspecific, ambiguous and unmatched classes of diseases from the list; hence, the catalog disease list was reduced to approximately 650 unique diseases.

3. Methodology

3.1. Construction of human associated networks

We have reconstructed two separate networks with the help of human gene-disease associations. The complete methodology and pipeline can be found in Fig. 1. In this study, we constructed two complementary networks, the Associated Gene Network (AGN) and the Associated Disease Network (ADN). In AGN (Additional file 1), nodes are genes and edges are their associated shared diseases, while in ADN (Additional file 2), nodes are diseases and edges are their associated shared genes. Both networks are based on an undirected weighted graph, where the number of associated shared genes or shared diseases is considered as the weight in the graph.

To reconstruct the Human Association Networks, we created Associated Gene and Disease Vectors. The Associated Disease Vector (Additional file 3) and Associated Gene Vector (Additional file 4) contain a list of shared genes for each disease and a list of shared diseases for each gene, respectively. To construct both vectors, the disease-gene relationships with p-values less than or equal to 0.01 were identified. Afterwards, the ADN and AGN networks were reconstructed using the vectors mentioned above by connecting diseases that shared genes and genes that are involved in common diseases, respectively. Fig. 2 shows the hub genes in ADN and Fig. 3 shows the hub genes in AGN. Genes with a higher degree than the average degree of all nodes are considered as hub nodes.

3.2. Evaluation of human association networks

Human Complex Network evaluation was performed using two approaches based on statistical and modular analysis, as well as with biological approaches.

I) Statistical and Modular Analysis

We have analyzed AGN and ADN topologically through the reconstruction of 1000 random null model networks. In these networks, we kept the degree constant while constructing random edges between nodes. We used assortativity as a reliable measure to qualify the results. We found the propensity of a node with a comparable degree size with a connecting edge. This measure finds the possibility of a connection between nodes with high degrees and nodes with low degrees and vice versa. Theoretically, r is considered to be the assortativity of a network by calculating the Pearson correlation coefficient of the degrees of the network on both sides of an edge over the set of all the edges [21,22]. In this study, we used assortativity as a topological network property to measure nodes with similar degree sizes with a connecting edge.

We used the total number of degree differences that are smaller or greater than a reference degree and then divided the acquired value by the number of random networks (1000) to obtain a two-tailed p-value as follows:

$$p - value = \frac{number\ of\ |a_i| > |a|}{number\ of\ random\ networks}$$

where $|a|$ denotes the absolute value of a, a_i is the assortativity of a random network i, and a is the assortativity for the real network (AGN/ADN). We used the false discovery rate (FDR) to rectify possible errors made due to multiple comparisons of p-values [23]. The p-values were considered significant at a p-value less than 0.05. The statistical analyses were performed using MATLAB R2016a and R (3.3.1).

Additionally, network modular analysis, which identifies different network modules based on the degree distribution and network connectivity, can be used as a validation measure. In this study, we identified different network clusters that are specified as highly connected subnetworks in both random null model networks and reference networks.

In this study, we used MCODE [24] in Cytoscape. We used the haircut method with a max depth of 100 nodes and a k-core cut-off of two. We counted the number of network modules revealed by both methods. As the number of network clusters in the random model networks were extensively higher than that of specified biological networks such as AGN/ADN, demonstrating that we can validate our analysis in this manner.

II) Biological Evidence

Apart from statistical and modular analysis, any systems biology approach needs to be validated through biological evidence to verify its applicability and functionality in biological and medical methods. In this study, the human association networks needed to be validated biologically to highlight the productivity and proficiency of the PheWAS studies when revealing novel findings.

We analyzed the Associated Gene Network and Associated Disease Network separately by different methods. The databases for the identification of gene-gene associations are based on knowledge from physical interaction databases such as MINT [25], HPRD [26], BIND [27], DIP [28], BioGRID [29], KEGG [30], Reactome [31], IntAct [32], EcoCyc [33], NCI-Nature Pathway Interaction Database [34] and Gene Ontology (GO) [35]. Cooccurring gene names were obtained by parsing scientific texts from SGD [36], OMIM [37], FlyBase [38] and PubMed. These operations have been analyzed using the STRING [39] tool.

The Associated Disease Network specified disease-disease associations through associated shared genes retrieved from the PheWAS

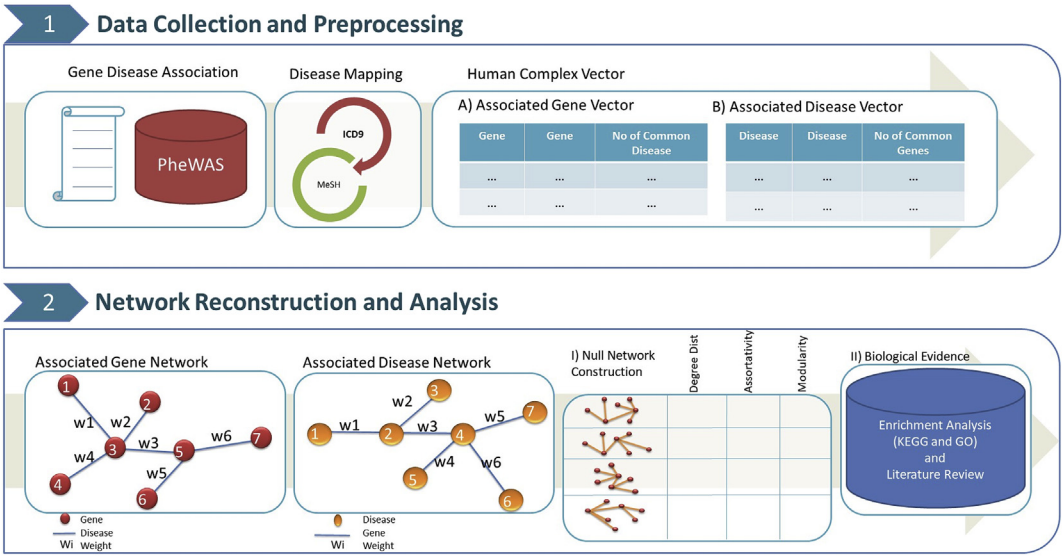

Fig. 1. Two step pipeline algorithm and methodology for the experiment. Preprocessing: We prepared data the based on a PheWAS catalog by data retrieval, mapped the ICD9 data with the MeSH database and then prepared the Human Complex Vectors. Network Reconstruction and Analysis: Reconstruction of AGN and ADN and their analysis with 1000 random null networks based on statistical measures and the examination of biological evidence.

catalogs. ADN can be used as a reference material for the identification of various disease-disease correlations. This study revealed the disease-disease network from other perspectives; hence, the biological validation of ADN. The modular analysis was prepared by MCODE [24] in Cytoscape. We used the haircut method with a max depth of 100 nodes and a k-core cut-off of two to identify ADN clusters. Additionally, for the biological validation of the ADN, some disease-disease relationships were selected as case studies. We tried to validate these relationships based on the literature mining using the PubMed database.

It is well understood that systematic analysis of genes and diseases is based on group-level data. Apart from that, the nature of the selected data was also from genetic population studies and not individual level data. To overcome the potential limitations of the analyses and interpretations, which may lead to ecologic bias, we have investigated the physical connectivity of genes and diseases by comparing AGN nodes and edges (genes and diseases, respectively) with a protein-protein interaction (PPI) network. In this study, we mapped the AGN network to a string database. To avoid any misplacement of interconnections, we examined the intersection of the mapped network with the AGN network; thus, we have nodes and edges present in both networks for further comparison.

4. Results

The result of our study is the construction of a human complex network from two different viewpoints. The Associated Gene Network (AGN) and Associated Disease Network (ADN) are the outcome of our analysis. These networks can be used as new references in systems biology and bioinformatics for finding novel gene/disease associations for any targeted phenotypic trait. Indeed, gene-gene and disease-disease associations are of great importance due to the ability to potentially identify drugs. In fact, drugs can be extracted from available drug compounds with the help of drug-repurposing methodologies. The results of each network are presented and discussed below.

4.1. Associated Gene Network

The AGN consists of 1769 genes connected to one another through 503 associative diseases. We acquired a larger negative assortativity for AGN compare with the randomized networks ($a = -0.035$ and $a = -0.143$ for AGN and randomized networks, respectively, p-value < 0.05). The comparative analysis of the random networks versus the specialized biological network showed that number of modules in any random network is comparatively higher than the specialized network. In this study, we observed the network clustering of AGN to be 18 network modules, which is comparatively much lesser than any random network of the same size.

Once the statistical analysis of the human association networks has been completed, the network is explored and analyzed biologically. As any complex diverse biological network such as AGN cannot be analyzed as a whole, we have selected a few case studies. These are selected by disease name, and accordingly, all genes associated with the selected disease have been extracted for further analysis. As the network is diverse and populated: thus, we have selected three different case studies, including colorectal cancer, skin cancer and type 2 diabetes. The number of associated genes in colorectal cancer, skin cancer and type 2 diabetes are 73, 114 and 73, respectively.

The contribution of many of the associated genes found in the AGN network for each phenotypic trait is of great importance. We have observed that researchers have previously reported many associated genes revealed by PheWAS studies. The text mining results for all the case studies are described in Table 1. As common genes found by the PheWAS-driven networks (AGN) and previous studies show the reliability of our approach, as the novel genes associated with any phenotypic trait are significant. However, the associations may not be supported by common biological evidence, such as protein-protein interactions, coexpression, or any other biological experiments, as these associations are caused by genetic variations and a single allele is responsible for their existence.

Apart from the above analysis, we analyzed our associated gene sets using KEGG pathway analysis. For this analysis, manually curated metabolic and signaling pathways were imported from KEGG (July 2016). We found that many associated genes in the AGN are present in different KEGG pathways. The results of this analysis are shown in Table 2. We observed that related pathway neighbors and subunits from the same enzyme/complex are effective and can lead to novel association links.

Gene Ontology analysis can impact the biological validation of the approach. The presence of highly connected associated genes can be reasonable evidence if there are one or many GO terms shared among them. Groups of skin cancer-associated genes were categorized as quaternary ammonium group transmembrane transporter activity (GO:

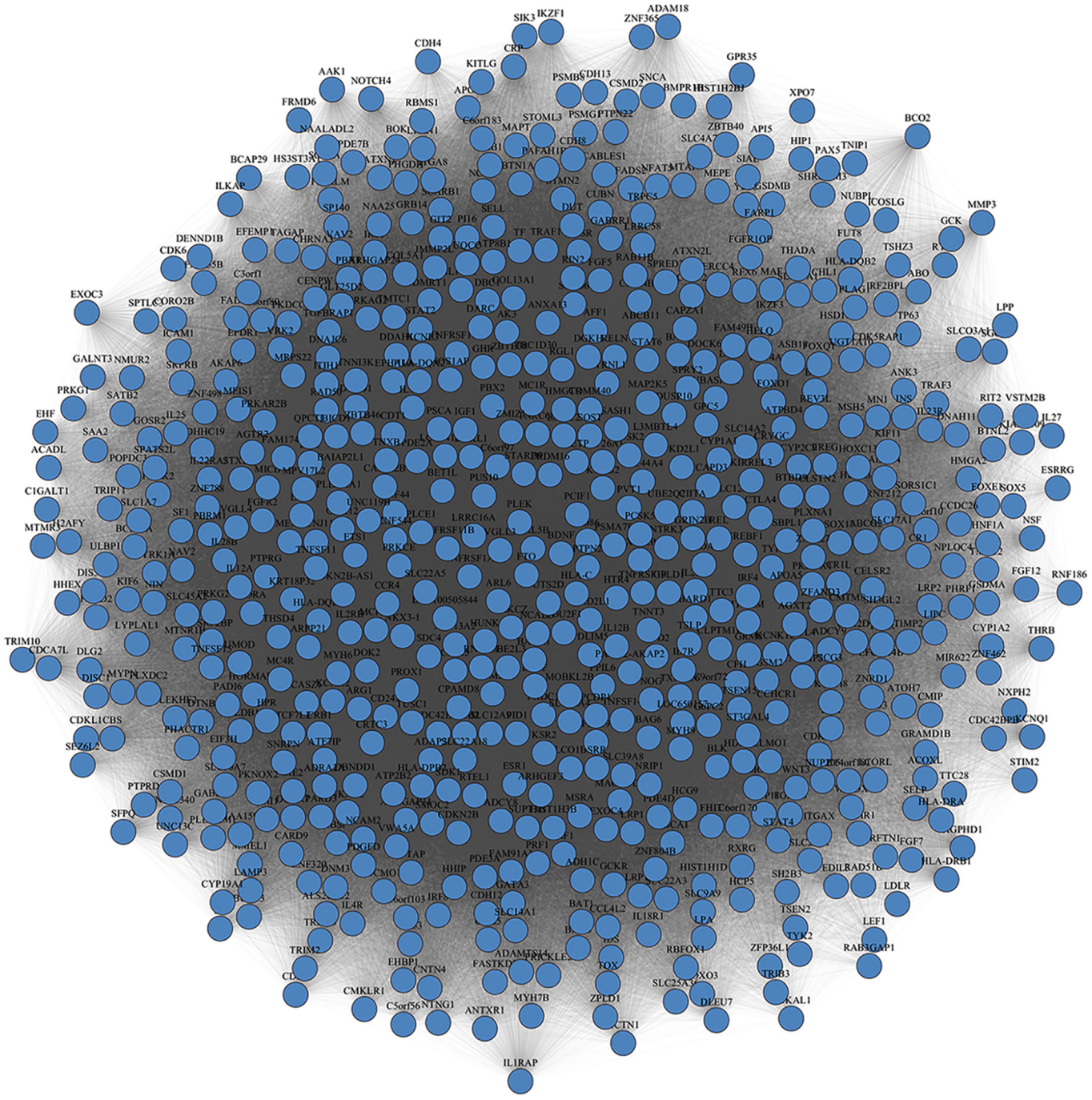

Fig. 2. Associated Gene Network. Nodes represent genes and edges represent diseases shared by common genes.

0015651) in molecular function [102]. Type 2 diabetes also shares a few genes categorized as MHC protein complex in cellular components with GO term GO:0042611 [103,104]. Finally, two cellular components share a few associated genes in colorectal cancer, including membrane (GO:0098552), which is an active cellular component in colorectal cancer [105], and cell surface (GO:0009986) [106–108].

4.2. Associated Disease Network

The Associated Disease Network (ADN) is a complementary network to AGN reconstructed from the interpretation of a PheWAS association matrix. The network shows disease-disease associations; thus, each node in the network represents a disease and each edge represents the number of shared genes associated with any two diseases. The ADN has 503 nodes and approximately 50000 connecting edges. We observed that skin-related diseases, such as psoriasis, cellulitis and skin cancer as well as types 2, 3, 4, and 6 diabetes, have the most connections with other diseases, which shows their centrality and complexity among all diseases.

Similar to the AGN, we analyzed the ADN statistically. We observed a negative assortativity for ADN that is larger than the randomized networks ($a = -0.036$ and $a = -0.142$ for ADN and the randomized networks, respectively; p-value < 0.05). We clustered ADN using the MCODE algorithm and achieved five network modules with node sizes of 166, 148, 91, 66 and 23. There are nine diseases that have not been coupled with any of the modules. Comparison of ADN to the randomized network shows that the number of network modules in ADN is much lower than that of the random network; hence, the network is

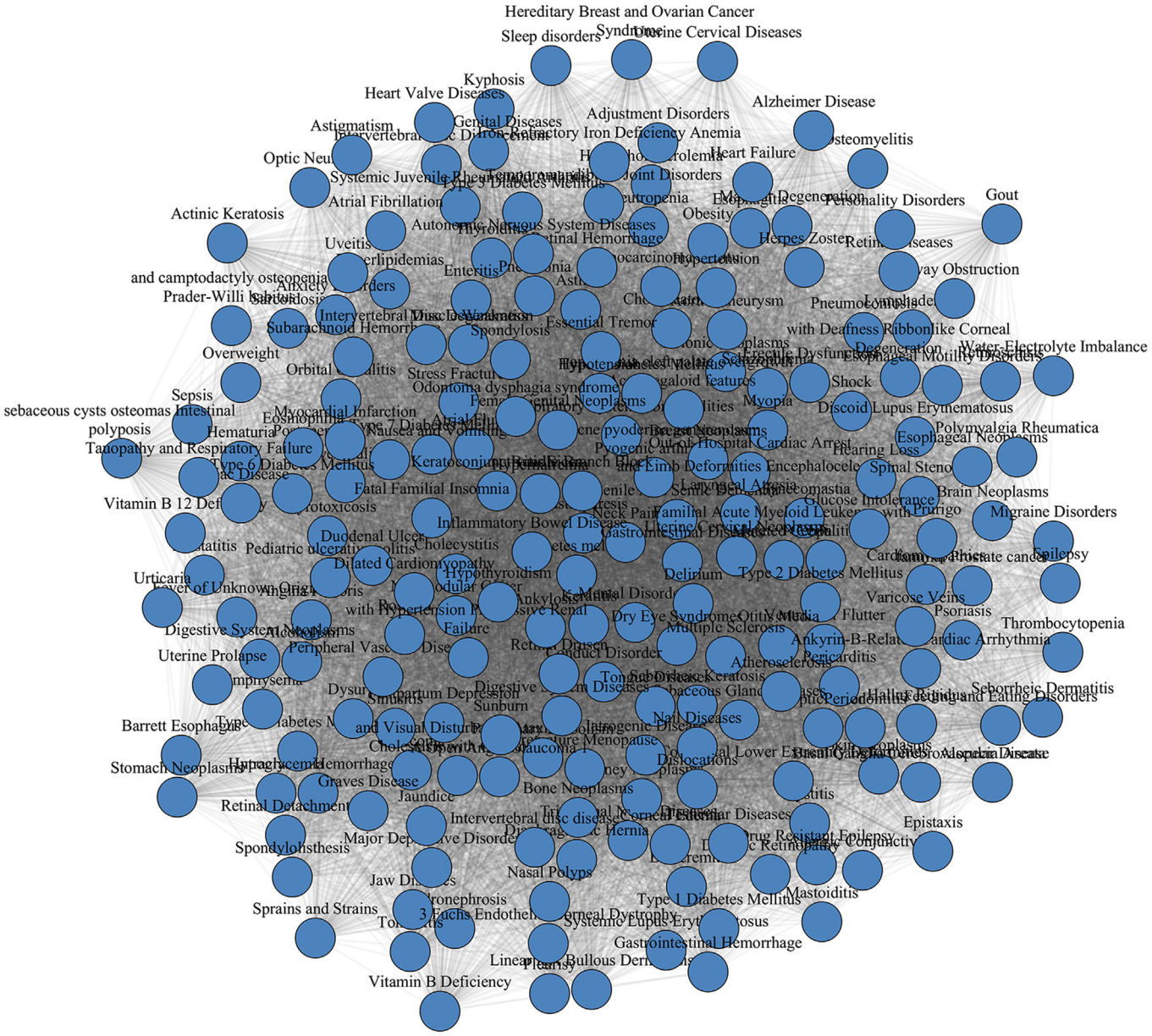

Fig. 3. Associated Disease Network. Nodes represent diseases and edges represent genes shared by common diseases.

significant.

For the biological validation of ADN, we have chosen a few diseases among the modules with higher numbers of associated genes. These diseases (analyzed in PubMed) were examined to validate their associations as a reference for associations reported in previous studies. The literature firmly supports our findings for the ADN network. Table 3 shows a few examples from our analysis.

Finally, we evaluated our findings, which are group level data that include unique individual level data. This analysis used a protein-protein interaction network from a string database as an individual level data and compared it with the AGN network. Fig. 4 shows the intersection network, which has been reconstructed from the AGN network and AGN-mapped PPI network as described earlier. This network contains 958 nodes interacting with 1884 edges. This indicates our group level data can also be justified with the help of PPI physical connections.

5. Discussion

Systematic analysis of PheWAS associations has many advantages and can be used as a reference to investigate and analyze genes and diseases. As the given approach can support researchers realizing the systematic associations of genes and diseases, many ignorant biological questions can be answered here. Identifying constitutive and responsive genes for specific diseases is always a challenge in the scientific world. Therefore, analysis of different allelic mutants and comorbidities, which has many affiliate reasons including environmental effects, shared biological mechanisms and phenotype similarity, through the help of PheWAS studies can begin a new era when identifying constitutive and responsive genes. Furthermore, these studies can help scientists find possible similarities in diseases due to analogous genes.

Disease-related interactome networks are a biological source for the reconstruction of metabolic, protein-protein interaction, gene regulatory, transcriptional profiling, phenotypic profiling, genetic interaction and integrating interactome networks [126]. The AGN and ADN networks are phenotypic profiling networks. Although all human

Table 1
Case study for AGN genes appearing in different diseases based on text mining and biological evidence analysis. Genes present in each row show evidence of multiple genes present in different studies, which were examined in the AGN network.

Disease Name	Gene Name	Reference
Colorectal Cancer	CD62L, CD62-P, CD56, CD95, CD178, CD95L, Fas-ligand, FasL, CD58, LFA-3, HLA-A	[40]
	Acid sphingomyelinase activity, sphingomyelinase, Acid sphingomyelinase activity, sphingomyelinase, FADD, APO-1, Apo-1, Apo1, CD95, FLICE, aspartate proteases, effector caspases, initiator caspase, initiator caspases, FasL, aspartate proteases	[41]
	phospholipase C, sphingomyelinase, phospholipase C, sphingomyelinase, FADD, APO-1, CD95, TNFRSF6, Caspase activity, caspase, caspases, effector caspases, FasL, Caspase activity	[42]
	Tris, MHC class I molecule, MHC class I molecule, Fas ligand	[43]
	Phospholipase A2, IP3 receptor, APO1, CD95, caspase	[44]
	FADD, APO-1, CD95, FLICE, calpain, calpains, executioner caspases, initiator caspases, FasL, PCNA, proliferating cell nuclear antigen, calpain	[45]
	L-selectin, F4/80, L-selectin, I-Ab	[46]
	phospholipases, CD56, elastase, elastases, FasL, elastase	[47]
	CD62L, Tris, MF23, IAb, MHC I	[48]
	HLA-DRB1, CD58, HLA-DQA1	[49]
	MYH11, CD56 HLA-DRB1, HLA-DRB1, ZNF157	[50]
Diabetes Type 2	MTNR1B, Melatonin receptor 1B, PROX1, CDKAL1, C2CD4B, SLC30A8, zinc transporter 8, FTO, TCF7L2	[51]
	MTNR1B, GCKR, glucokinase regulatory protein, CDKAL1, HHEX, FTO, TCF7L2	[52]
	CDKAL1, HHEX, JAZF1, SLC30A8, ZnT-8, FTO, fat mass and obesity associated, TCF7L2, transcription factor 7-like 2	[53]
	CDKAL1, HHEX, JAZF1, SLC30A8, FTO, TCF7L2	[54]
	CDKAL1, HHEX, JAZF1, C2CD4B, SLC30A8, FTO, TCF7L2	[55]
	CDKAL1, HHEX, JAZF1, SLC30A8, FTO, TCF7L2	[56]
	MTNR1B, HHEX, DUSP-9, DUSP9, MKP-4, SLC30A8, FTO, TCF7L2	[57]
	MTNR1B, CDKAL1, HHEX, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DQB1, SLC30A8, ZnT8, zinc transporter 8, TCF7L2	[58]
	MTNR1B, CDKAL1, HHEX, DUSP-9, DUSP9, Dusp9, MKP-4, MKP4, Mkp4, mitogen-activated protein kinase phosphatase 4, C2CD4B, SLC30A8, TCF7L2, transcription factor 7-like 2	[59]
	hepatocyte nuclear factor 1alpha, KIF11, CDKAL1, HHEX, Hex, haematopoietically expressed homeobox, JAZF1, SLC30A8, TCF7L2	[60]
	MTNR1B, CDKAL1, HHEX, C2CD4B, SLC30A8, TCF7L2, transcription factor 7-like 2	[61]
	MTNR1B, hepatocyte nuclear factor-1alpha, PROX1, GCKR, glucokinase regulatory protein, CDKAL1, JAZF1, SLC30A8, TCF7L2	[62]
	MTNR1B, Melatonin receptor 1B, melatonin receptor 1B, HNF-1 alpha, HNF1 homeobox A, HNF1A, HNF1alpha, MODY3, PROX1, Prospero homeobox 1, prosperohomeobox 1, glucokinase regulatory protein, CDK5 regulatory subunit associated protein 1-like 1, CDKAL1, JAZF zinc finger 1, JAZF1, C2 calcium-dependent domain containing 4B, C2 calcium-dependent domain containing 4B, C2CD4B, SLC30A8, TCF-4, TCF7L2, Transcription factor 7-like 2, transcription factor 7-like 2, transcription factor-7-like 2	[63]
	Melatonin receptor 1B, PROX1, Prospero homeobox 1, glucokinase regulatory protein, CDK5 regulatory subunit-associated protein 1-like 1, CDKAL1, cdkal1, C2 calcium-dependent domain containing 4B, SLC30A8, Solute carrier family 30 member 8, ZNT8, zinc transporter 8, TCF7L2, Transcription factor 7-like 2	[64]
	MTNR1B, KIF11, CDK5 regulatory subunit-associated protein 1-like 1, CDKAL1, JAZF1, C2CD4B, SLC30A8, ZnT8, TCF7L2, Transcription factor 7-like 2, transcription factor 7-like 2	[65]
	MTNR1B, PROX1, CDKAL1, C2CD4B, SLC30A8, TCF7L2	[66,67]
	MTNR1B, CDKAL1, HHEX, IRS1, C2CD4A, C2CD4B, HHEX, SLC30A8, TCF7L2	[59]
	MTNR1B, CDKAL1, HHEX, JAZF1, HHEX, SLC30A8, FTO, TCF7L2	[68]
Skin Cancer	EXOC2, exocyst complex component 2, SLC45A2, IRF4, interferon regulatory factor 4, MC1R, melanocortin receptor 1	[69]
	ST3GAL4, SEMA6A, IRF4, HLA-DRA, MC1R	[70]
	EXOC2, MATP, SLC45A2, IRF4, MC1R, Melanocortin 1 receptor	[71]
	TOX3, LRP5, MATP, SLC45A2, membrane-associated transporter protein, solute carrier family 45 member 2, LSP1, MC1R, melanocortin 1 receptor, melanocortin-1 receptor	[72]
	MATP, SLC45A2, CDK10, IRF4, MC1R, Melanocortin 1 receptor, Melanocortin Receptor-1, Melanocortin-1 receptor, melanocortin 1 receptor, melanocortin-1-receptor	[73]
	tyrosinase, c Ab, I Ab	[74]

Table 2
Case study for selected AGN genes based on KEGG pathways in different diseases. Each row represents a KEGG pathway related to the mentioned disease and number of genes present in the AGN network with a p-value less than 1e-6.

Disease Name	Pathway Name	No of Gene	Reference
Diabetes Type 2	HTLV-I infection	6	[75,76]
	Cell adhesion molecules (CAMs)	5	[77–80]
	Epstein-Barr virus infection	5	[81,82]
	Staphylococcus aureus infection	5	[83,84]
	Toxoplasmosis	4	[85–87]
Skin Cancer	Melanogenesis	5	[88,89]
	HTLV-I infection	7	[90,91]
Colorectal Cancer	Cell adhesion molecules (CAMs)	9	[92,93]
	Graft-versus-host disease	5	[94,95]
	Type I diabetes mellitus	5	[96–99]
	Autoimmune thyroid disease	5	[100,101]

disease networks are not reliable and unique sources for reconstruction [127–129], they can be compared and contrasted with each other. It's highly advised to compare and contrast networks with similar sources of reconstruction. Genome Wide Association Studies (GWAS) have similar allelic variant association contents for diseases and can be compared to the AGN and ADN networks. It has been observed that complex human disease gene networks reconstructed from the allelic contents of GWAS are comparable with our networks [130]. As complex human disease gene networks constructed from 54 diseases and 349 genes, the complexity of our networks is comparatively less complicated and easy to interpret. The nature of the nodes and interactions in complex human disease gene networks are highly similar and comparable.

AGN network reconstruction can help us to identify constitutive and responsive genes for various diseases. Statistical analysis of the AGN network showed common genes shared by most of the diseases. This investigation showed that the common shared genes in the AGN are normal, nonresponsive genes that are semiresponsive in most biological pathways and that their effective roles cannot be observed in specific diseases. In this regard, constitutive and responsive genes can be found by analyzing more strategic nodes in the AGN network with the help of their incidence, position and degree distributions.

As constitutive and responsive genes can be identified as potentially well-established drug targets, this analysis can be used to identify various drug targets for phenotypic traits. Hence, drug repurposing can

Table 3
Disease associations from the ADN network using PubMed. The highly correlated diseases based on ADN analysis using different studies.

Diseases-Disease relationship from ADN	Description	References
Type 1 and Type 2 Diabetes Mellitus, Systemic Lupus Erythematosus (SLE)	Diabetes mellitus complicating systemic lupus erythematosus	[109]
Type 1 Diabetes Mellitus (T1D), Systemic Lupus Erythematosus (SLE), Rheumatoid Arthritis (RA)	The BTNL2 polymorphism is associated with T1D, RA, and SLE because of its strong linkage disequilibrium with predisposing HLA DQB1-DRB1 haplotypes in Caucasian populations.	[110]
Celiac Disease (CD), Systemic Lupus Erythematosus (SLE)	Individuals with CD were at a 3-fold increased risk of SLE compared to the general population.	[111]
	underlying mechanism for the positive association between celiac disease and systemic lupus erythematosus: the role of interleukin 21	[112]
	CD and SLE are autoimmune diseases and share HLAB8 and DR3 histocompatibility-antigens.	[113]
	This study suggests that SLE occurs far more frequently in biopsy-defined celiac disease than is currently appreciated	[114]
Hypoglycemia, type 2 diabetes mellitus	In patients aged ≥65 years with type 2 diabetes, higher glucose variability and lower average glucose levels indicate a greater hypoglycemia risk. It is therefore necessary to ensure comprehensive blood glucose control in these patients to prevent hypoglycemia.	[115]
	The reported prevalence of hypoglycemia among type 2 diabetes patients is quite high.	[116]
	Association between hypoglycemia and fall-related events in type 2 diabetes mellitus	[117]
Psoriasis, skin cancer	Personal history of psoriasis may be associated with an increased risk of SCC	[118]
	We observed a modestly increased risk of melanoma and nonmelanoma skin cancers in patients with mild psoriasis, whereas patients with severe psoriasis and psoriatic arthritis had an increased risk of nonmelanoma skin cancer but not melanoma.	[119]
	Psoriasis carries an elevated risk of nonmelanoma skin cancer and lymphoma. This effect is modified by the severity of psoriasis, age, gender, and geographic location.	[120]
Diabetes Mellitus Type 2, Asthma	The prevalence of asthma is significantly higher in hospitalized patients with type II DM independent of other comorbid conditions.	[121]
	The authors found that the retrospective diagnosis of asthma was related to a significantly greater risk of DM compared with controls	[122]
Psoriasis, Celiac Disease	Our meta-analysis demonstrated an approximately 3-fold increased risk among patients with psoriasis	[123]
	Association between celiac disease and psoriasis	[124]
	High prevalence of celiac disease in psoriasis	[125]

be used as an alternative path to find therapeutic approaches with the help of potential target genes.

ADN network analysis also showed various disease correlations and similarity. The similarity of diseases, which can be found by analyzing the nodes and edges, demonstrate the overlapping responsive or semi-responsive genes between diseases. Comorbidity and phenotype similarities can be easily analyzed with the ADN network. This can be used to examine drug efficacy in different diseases with genetic similarities, side effects and drug repurposing.

AGN and ADN network analysis is based on group-level data and does not apply to individual-level data. We do not suggest this analysis for hypothesis testing due to limitations of the analysis and interpretation based on ecologic bias. There are various approaches used by individual-based data, which can support our findings. However, there is no individual level data analysis included here, indicating all kinds of disease or gene associations, but these associations can be validated with the help of these approaches. The PPI-mapped AGN network analysis showed over 950 nodes (genes), with large numbers of interaction. The association of AGN nodes are based on diseases and may not have direct physical interactions. Additionally, AGN network analysis has been accomplished by GO and KEGG enrichment (Table 3) of associated genes in specific diseases investigated in the case studies. The approach shows supporting evidence from the biological processes, molecular function, and cellular component categories associated with GO analysis and metabolic pathways associated with KEGG analysis. Other supporting evidence is found in Tables 1 and 2 Thus, this approach is recommended for generating hypotheses in the biomedical sciences.

6. Conclusions

PheWAS is a well-established methodology for the analysis of genetic variations and generates huge amounts of integral information that can be analyzed and investigated systematically. Systems biology approaches can be applied to PheWAS information to build Human Association Networks. The Associated Gene Network and the Associated Disease Network are two complementary networks that were reconstructed from PheWAS associations. AGN showed novel association of different genes found in biological experiments and/or databases. These genetic variations were highlighted due to shared diseases but occur due to various scientific reasons, such as the environment, population, life style, etc. These networks can be used to explore medical and biological evidence from a different perspective. Apart from that, the ADN can help identify new disease associations based on common shared genes. The identification of potential drug compounds for any disease can use this systematic approach. This network conveys vast amounts of information for many purposes such as drug repurposing, disease analysis, identification of interconnections and relationships between different diseases and many more.

Acknowledgements

We would like to thank Dr. Hossein Seidkhani for his valuable comments on the statistical analysis.

Abbreviations

AGN	Associated gene network
ADN	Associated disease network
PheWAS	Phenome Wide Association Studies
GWAS	Genome Wide Association Studies
ICD9	International Classification of Diseases version 9
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
EMR	electronic medical records
MESH	Medical Subject Headings

Ethics approval and consent to participate

Not Applicable.

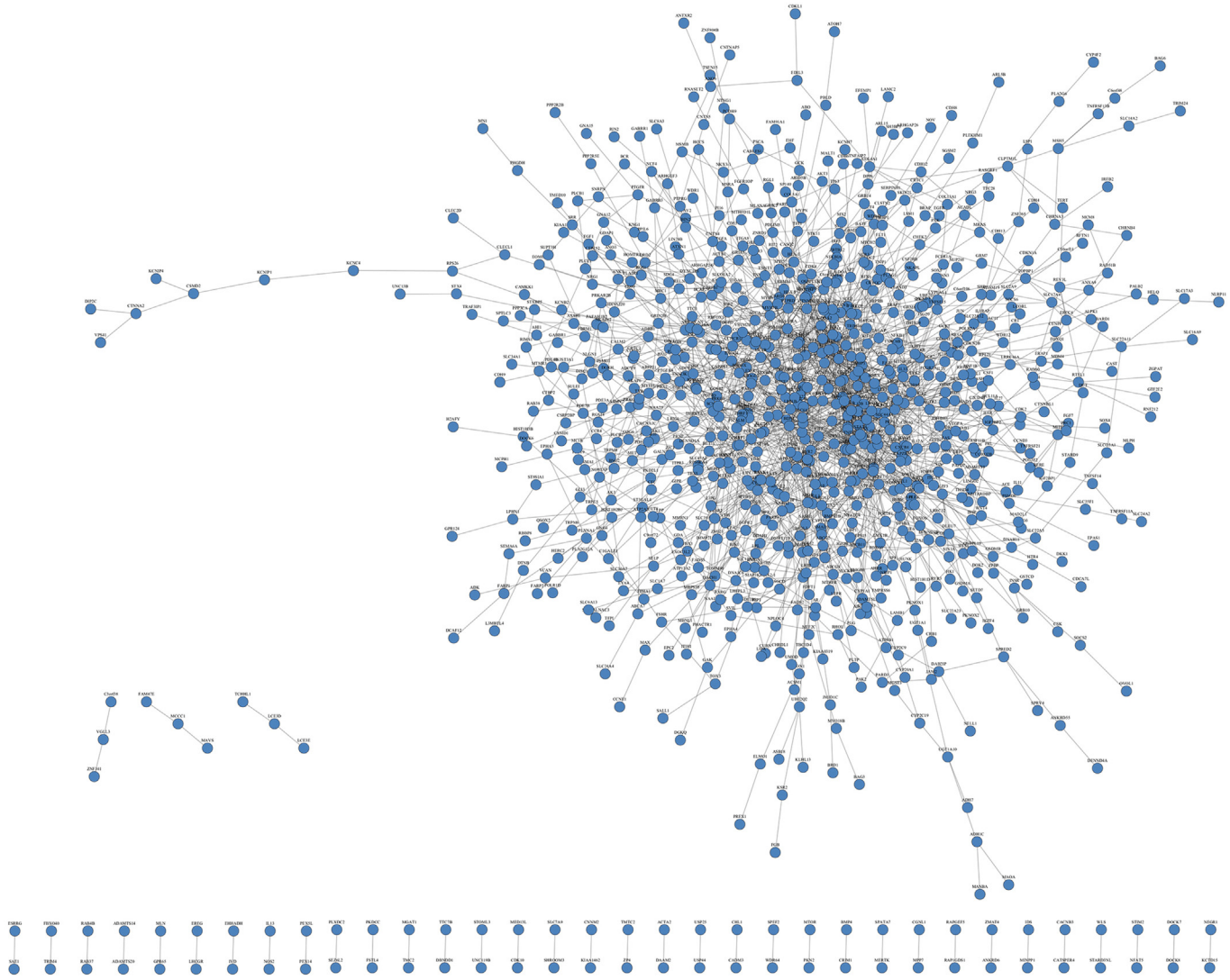

Fig. 4. The intersection network between the AGN and reconstructed PPI-mapped AGN network. This network has common nodes and edges in both the AGN and PPI-derived networks from the string database. The connecting edges from the AGN network were compared with PPI physical interactions as individual level data.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that they have no competing interests.

Authors' contribution

AK conceived and designed the experiments, analyzed the data, and wrote the manuscript; MK supervised the network reconstruction and analysis; BG, AM, BJ revised the manuscript and designed the experiments; and all authors have read and approved the manuscript.

Additional files

Additional file 1: List of final nodes used to construct the Associated Gene Network. The first two columns are the gene names, which are connected by an edge, and the third column is the number of shared diseases.

Additional file 2: List of final nodes used to construct the Associated Disease Network. The first two columns are the disease names, which are connected by an edge, and third column is the number of shared genes.

Additional file 3: Disease names and list of shared genes used as a set for the Associated Disease Vector.

Additional file 4: Gene names and list of shared diseases used as a set for the Associated Gene Vector.

References

[1] M. Rastegar-Mojarad, Z. Ye, J.M. Kolesar, S.J. Hebbring, S.M. Lin, Opportunities for drug repositioning from phenome-wide association studies, Nat. Biotechnol. 33 (2015) 342–345.

[2] J.C. Denny, L. Bastarache, D.M. Roden, Phenome-wide association studies as a tool to advance precision medicine, Annu. Rev. Genom. Hum. Genet. 17 (2014) 1–21.

[3] J.C. Denny, L. Bastarache, D.M. Roden, Phenome-wide association studies as a tool to advance precision medicine, Annu. Rev. Genom. Hum. Genet. 17 (2016) 353–373.

[4] J.C. Denny, M.D. Ritchie, M.A. Basford, J.M. Pulley, L. Bastarache, K. Brown-Gentry, D. Wang, D.R. Masys, D.M. Roden, D.C. Crawford, PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations, Bioinformatics 26 (2010) 1205–1210.
[5] S.J. Hebbring, The challenges, advantages and future of phenome-wide association studies, Immunology 141 (2014) 157–165.
[6] S.A. Pendergrass, K. Brown-Gentry, S.M. Dudek, E.S. Torstenson, J.L. Ambite, C.L. Avery, S. Buyske, C. Cai, M.D. Fesinmeyer, C. Haiman, G. Heiss, L.A. Hindorff, C.N. Hsu, R.D. Jackson, C. Kooperberg, L. Le Marchand, Y. Lin, T.C. Matise, L. Moreland, K. Monroe, A.P. Reiner, R. Wallace, L.R. Wilkens, D.C. Crawford, M.D. Ritchie, The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery, Genet. Epidemiol. 35 (2011) 410–422.
[7] J.C. Denny, L. Bastarache, M.D. Ritchie, R.J. Carroll, R. Zink, J.D. Mosley, J.R. Field, J.M. Pulley, A.H. Ramirez, E. Bowton, M.A. Basford, D.S. Carrell, P.L. Peissig, A.N. Kho, J.A. Pacheco, L.V. Rasmussen, D.R. Crosslin, P.K. Crane, J. Pathak, S.J. Bielinski, S.A. Pendergrass, H. Xu, L.A. Hindorff, R. Li, T.A. Manolio, C.G. Chute, R.L. Chisholm, E.B. Larson, G.P. Jarvik, M.H. Brilliant, C.A. McCarty, I.J. Kullo, J.L. Haines, D.C. Crawford, D.R. Masys, D.M. Roden, Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data, Nat. Biotechnol. 31 (2013) 1102–1110.
[8] R.M. Plenge, E.M. Scolnick, D. Altshuler, Validating therapeutic targets through human genetics, Nat. Rev. Drug Discov. 12 (2013) 581–594.
[9] R. Xu, Q. Wang, A genomics-based systems approach towards drug repositioning for rheumatoid arthritis, BMC Genomics 17 (2016) 518.
[10] J. Zhang, K. Jiang, L. Lv, H. Wang, Z. Shen, Z. Gao, B. Wang, Y. Yang, Y. Ye, S. Wang, Use of genome-wide association studies for cancer research and drug repositioning, PLoS One 10 (2015) e0116477.
[11] H. Wang, Q. Gu, J. Wei, Z. Cao, Q. Liu, Mining drug-disease relationships as a complement to medical genetics-based drug repositioning: where a recommendation system meets genome-wide association studies, Clin. Pharmacol. Ther. 97 (2015) 451–454.
[12] P. Sanseau, P. Agarwal, M.R. Barnes, T. Pastinen, B.J. Richards, L.R. Cardon, V. Mooser, Use of genome-wide association studies for drug repositioning, Nat. Biotechnol. 30 (2012) 317–320.
[13] R. Xu, Q. Wang, PhenoPredict: a disease phenome-wide drug repositioning approach towards schizophrenia drug discovery, J. Biomed. Inform. 56 (2015) 348–355.
[14] M. Rastegar-Mojarad, H. Liu, P. Nambisan, Using social media data to identify potential candidates for drug repurposing: a feasibility study, JMIR Res. Protoc. 5 (2016).
[15] S.A. Pendergrass, K. Brown-Gentry, S. Dudek, A. Frase, E.S. Torstenson, R. Goodloe, J.L. Ambite, C.L. Avery, S. Buyske, P. Buzkova, E. Deelman, M.D. Fesinmeyer, C.A. Haiman, G. Heiss, L.A. Hindorff, C.N. Hsu, R.D. Jackson, C. Kooperberg, L. Le Marchand, Y. Lin, T.C. Matise, K.R. Monroe, L. Moreland, S.L. Park, A. Reiner, R. Wallace, L.R. Wilkens, D.C. Crawford, M.D. Ritchie, Phenome-wide association study (PheWAS) for detection of pleiotropy within the population architecture using genomics and epidemiology (PAGE) network, PLoS Genet. 9 (2013) e1003087.
[16] S. Pendergrass, K. Brown-Gentry, S. Dudek, E. Torstenson, J. Ambite, C. Avery, S. Buyske, C. Cai, M. Fesinmeyer, C. Haiman, The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery, Genet. Epidemiol. 35 (2011) 410–422.
[17] J.C. Denny, M.D. Ritchie, M.A. Basford, J.M. Pulley, L. Bastarache, K. Brown-Gentry, D. Wang, D.R. Masys, D.M. Roden, D.C. Crawford, PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations, Bioinformatics 26 (2010) 1205–1210.
[18] J.C. Denny, L. Bastarache, M.D. Ritchie, R.J. Carroll, R. Zink, J.D. Mosley, J.R. Field, J.M. Pulley, A.H. Ramirez, E. Bowton, M.A. Basford, D.S. Carrell, P.L. Peissig, A.N. Kho, J.A. Pacheco, L.V. Rasmussen, D.R. Crosslin, P.K. Crane, J. Pathak, S.J. Bielinski, S.A. Pendergrass, H. Xu, L.A. Hindorff, R. Li, T.A. Manolio, C.G. Chute, R.L. Chisholm, E.B. Larson, G.P. Jarvik, M.H. Brilliant, C.A. McCarty, I.J. Kullo, J.L. Haines, D.C. Crawford, D.R. Masys, D.M. Roden, Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data, Nat. Biotechnol. 31 (2013) 1102–1111.
[19] D.A. Bello SM, T.C. Wiegers, M.E. Dolan, C. Smith, J. Richardson, J. Blake, J.T. Eppig, C.J. Mattingly, Waiting for a robust Disease Ontology: a merger of OMIM and MeSH as a practical interim solution, International Conference on Biomedical Ontology (ICBO), 2011, pp. 231–233.
[20] C.E. Lipscomb, Medical Subject Headings (MeSH), Bull. Med. Libr. Assoc. 88 (2000) 265–266.
[21] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (2002) 47.
[22] M.E. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256.
[23] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B (1995) 289–300.
[24] G.D. Bader, C.W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, BMC Bioinf. 4 (2003) 2.
[25] L. Licata, L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A.P. Nardozza, E. Santonico, L. Castagnoli, G. Cesareni, MINT, the molecular interaction database: 2012 update, Nucleic Acids Res. 40 (2012) D857–D861.
[26] T.S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D.S. Somanathan, A. Sebastian, S. Rani, S. Ray, C.J. Harrys Kishore, S. Kanth, M. Ahmed, M.K. Kashyap, R. Mohmood, Y.L. Ramachandra, V. Krishna, B.A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, A. Pandey, Human protein reference database–2009 update, Nucleic Acids Res. 37 (2009) D767–D772.
[27] G.D. Bader, BIND: the biomolecular interaction network database, Nucleic Acids Res. 31 (2003) 248–250.
[28] I. Xenarios, E. Fernandez, L. Salwinski, X.J. Duan, M.J. Thompson, E.M. Marcotte, D. Eisenberg, DIP: the database of interacting proteins: 2001 update, Nucleic Acids Res. 29 (2001) 239–241.
[29] A. Chatr-Aryamontri, B.J. Breitkreutz, R. Oughtred, L. Boucher, S. Heinicke, D. Chen, C. Stark, A. Breitkreutz, N. Kolas, L. O'Donnell, T. Reguly, J. Nixon, L. Ramage, A. Winter, A. Sellam, C. Chang, J. Hirschman, C. Theesfeld, J. Rust, M.S. Livstone, K. Dolinski, M. Tyers, The BioGRID interaction database: 2015 update, Nucleic Acids Res. 43 (2015) D470–D478.
[30] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs, Nucleic Acids Res. 45 (2017) D353–d361.
[31] D. Croft, G. O'Kelly, G. Wu, R. Haw, M. Gillespie, L. Matthews, M. Caudy, P. Garapati, G. Gopinath, B. Jassal, S. Jupe, I. Kalatskaya, S. Mahajan, B. May, N. Ndegwa, E. Schmidt, V. Shamovsky, C. Yung, E. Birney, H. Hermjakob, P. D'Eustachio, L. Stein, Reactome: a database of reactions, pathways and biological processes, Nucleic Acids Res. 39 (2011) D691–D697.
[32] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N.H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R.C. Lovering, B. Meldal, A.N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni, H. Hermjakob, The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases, Nucleic Acids Res. 42 (2014) D358–D363.
[33] I.M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muniz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A.G. Shearer, A. Mackie, I. Paulsen, R.P. Gunsalus, P.D. Karp, EcoCyc: a comprehensive database of Escherichia coli biology, Nucleic Acids Res. 39 (2011) D583–D590.
[34] K. Anthony, S. Krupa, J. Buchoff, M. Day, T. Hannay, C. Schaefer, The NCI-Nature Pathway Interaction Database: a cell signaling resource, Nature Precedings (2007) 153–158.
[35] M.A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G.M. Rubin, J.A. Blake, C. Bult, M. Dolan, H. Drabkin, J.T. Eppig, D.P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J.M. Cherry, K.R. Christie, M.C. Costanzo, S.S. Dwight, S. Engel, D.G. Fisk, J.E. Hirschman, E.L. Hong, R.S. Nash, A. Sethuraman, C.L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S.Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E.M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, R. White, The Gene Ontology (GO) database and informatics resource, Nucleic Acids Res. 32 (2004) D258–D261.
[36] J.M. Cherry, C. Adler, C. Ball, S.A. Chervitz, S.S. Dwight, E.T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng, D. Botstein, SGD: Saccharomyces genome database, Nucleic Acids Res. 26 (1998) 73–79.
[37] A. Hamosh, A.F. Scott, J. Amberger, D. Valle, V.A. McKusick, Online Mendelian inheritance in man (OMIM), Hum. Mutat. 15 (2000) 57.
[38] M.A. Crosby, J.L. Goodman, V.B. Strelets, P. Zhang, W.M. Gelbart, C. FlyBase, FlyBase: genomes by the dozen, Nucleic Acids Res. 35 (2007) D486–D491.
[39] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life, Nucleic Acids Res. 43 (2015) D447–D452.
[40] L. Belov, J. Zhou, R.I. Christopherson, Cell surface markers in colorectal cancer prognosis, Int. J. Mol. Sci. 12 (2011) 78–113.
[41] A. Apraiz, M.D. Boyano, A. Asumendi, Cell-centric view of apoptosis and apoptotic cell death-inducing antitumoral strategies, Cancers 3 (2011) 1042–1080.
[42] J. Hofmanova, N. Strakova, A.H. Vaculova, Z. Tylichova, B. Safarikova, B. Skender, A. Kozubik, Interaction of dietary fatty acids with tumour necrosis factor family cytokines during colon inflammation and cancer, Mediat. Inflamm. 2014 (2014) 848632.
[43] M. Wick, P. Dubey, H. Koeppen, C.T. Siegel, P.E. Fields, L. Chen, J.A. Bluestone, H. Schreiber, Antigenic cancer cells grow progressively in immune hosts without evidence for T cell exhaustion or systemic anergy, J. Exp. Med. 186 (1997) 229–238.
[44] C. Huang, C. Freter, Lipid metabolism, apoptosis and cancer therapy, Int. J. Mol. Sci. 16 (2015) 924–949.
[45] S. Bhattacharya, M.K. Ghosh, Cell death and deubiquitinases: perspectives in cancer, BioMed Res. Int. 2014 (2014) 435197.
[46] W. Jing, J.A. Gershan, J. Weber, D. Tlomak, L. McOlash, C. Sabatos-Peyton, B.D. Johnson, Combined immune checkpoint protein blockade and low dose whole body irradiation as immunotherapy for myeloma, J. Immunother. Canc. 3 (2015) 2.
[47] A. Benito-Martin, A. Di Giannatale, S. Ceder, H. Peinado, The new deal: a potential role for secreted vesicles in innate immunity and tumor progression, Front. Immunol. 6 (2015) 66.
[48] N.R. Mahadevan, V. Anufreichik, J.J. Rodvold, K.T. Chiu, H. Sepulveda, M. Zanetti, Cell-extrinsic effects of tumor ER stress imprint myeloid dendritic cells

and impair CD8(+) T cell priming, PLoS One 7 (2012) e51845.
[49] S. Falgreen, K. Dybkaer, K.H. Young, Z.Y. Xu-Monette, T.C. El-Galaly, M.B. Laursen, J.S. Bodker, M.K. Kjeldsen, A. Schmitz, M. Nyegaard, H.E. Johnsen, M. Bogsted, Predicting response to multidrug regimens in cancer patients using cell line experiments and regularised regression models, BMC Canc. 15 (2015) 235.
[50] E. Bonnet, M. Tatari, A. Joshi, T. Michoel, K. Marchal, G. Berx, Y. Van de Peer, Module network inference from a cancer gene expression data set identifies microRNA regulated modules, PLoS One 5 (2010) e10162.
[51] R.B. Prasad, L. Groop, Genetics of type 2 diabetes-pitfalls and possibilities, Genes 6 (2015) 87–123.
[52] X. Kong, X. Zhang, X. Xing, B. Zhang, J. Hong, W. Yang, The association of type 2 diabetes loci identified in genome-wide association studies with metabolic syndrome and its components in a Chinese population with type 2 diabetes, PLoS One 10 (2015) e0143607.
[53] N. Abdullah, J. Attia, C. Oldmeadow, R.J. Scott, E.G. Holliday, The architecture of risk for type 2 diabetes: understanding Asia in the context of global findings, Int. J. Endocrinol. 2014 (2014) 593982.
[54] R. Villegas, R.J. Goodloe, B.E. McClellan Jr., J. Boston, D.C. Crawford, Gene-carbohydrate and gene-fiber interactions and type 2 diabetes in diverse populations from the national health and nutrition examination surveys (NHANES) as part of the epidemiologic architecture for genes linked to environment (EAGLE) study, BMC Genet. 15 (2014) 69.
[55] Y.C. Chang, P.H. Liu, Y.H. Yu, S.S. Kuo, T.J. Chang, Y.D. Jiang, J.Y. Nong, J.J. Hwang, L.M. Chuang, Validation of type 2 diabetes risk variants identified by genome-wide association studies in Han Chinese population: a replication study and meta-analysis, PLoS One 9 (2014) e95045.
[56] C.H. Tam, J.S. Ho, Y. Wang, V.K. Lam, H.M. Lee, G. Jiang, E.S. Lau, A.P. Kong, X. Fan, J.L. Woo, S.K. Tsui, M.C. Ng, W.Y. So, J.C. Chan, R.C. Ma, Use of net reclassification improvement (NRI) method confirms the utility of combined genetic risk score to predict type 2 diabetes, PLoS One 8 (2013) e83093.
[57] R. Ribel-Madsen, M.F. Fraga, S. Jacobsen, J. Bork-Jensen, E. Lara, V. Calvanese, A.F. Fernandez, M. Friedrichsen, B.F. Vind, K. Hojlund, H. Beck-Nielsen, M. Esteller, A. Vaag, P. Poulsen, Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes, PLoS One 7 (2012) e51302.
[58] C. Winkler, J. Raab, H. Grallert, A.G. Ziegler, Lack of association of type 2 diabetes susceptibility genotypes and body weight on the development of islet autoimmunity and type 1 diabetes, PLoS One 7 (2012) e35410.
[59] H. Fukuda, M. Imamura, Y. Tanaka, M. Iwata, H. Hirose, K. Kaku, H. Maegawa, H. Watada, K. Tobe, A. Kashiwagi, R. Kawamori, S. Maeda, A single nucleotide polymorphism within DUSP9 is associated with susceptibility to type 2 diabetes in a Japanese population, PLoS One 7 (2012) e46263.
[60] Y. Qian, F. Lu, M. Dong, Y. Lin, H. Li, J. Chen, C. Shen, G. Jin, Z. Hu, H. Shen, Genetic variants of IDE-KIF11-HHEX at 10q23.33 associated with type 2 diabetes risk: a fine-mapping study in Chinese population, PLoS One 7 (2012) e35060.
[61] M. Iwata, S. Maeda, Y. Kamura, A. Takano, H. Kato, S. Murakami, K. Higuchi, A. Takahashi, H. Fujita, K. Hara, T. Kadowaki, K. Tobe, Genetic risk score constructed using 14 susceptibility alleles for type 2 diabetes is associated with the early onset of diabetes and may predict the future requirement of insulin injections among Japanese individuals, Diabetes Care 35 (2012) 1763–1770.
[62] A. Stancakova, J. Paananen, P. Soininen, A.J. Kangas, L.L. Bonnycastle, M.A. Morken, F.S. Collins, A.U. Jackson, M.L. Boehnke, J. Kuusisto, M. Ala-Korpela, M. Laakso, Effects of 34 risk loci for type 2 diabetes or hyperglycemia on lipoprotein subclasses and their composition in 6,580 nondiabetic Finnish men, Diabetes 60 (2011) 1608–1616.
[63] E. Ahlqvist, T.S. Ahluwalia, L. Groop, Genetics of type 2 diabetes, Clin. Chem. 57 (2011) 241–254.
[64] R.D. Cox, C.D. Church, Mouse models and the interpretation of human GWAS in type 2 diabetes and obesity, Dis. Model. Mech. 4 (2011) 155–164.
[65] K.S. Park, The search for genetic risk factors of type 2 diabetes mellitus, Diabetes Metab. J. 35 (2011) 12–22.
[66] T. Ohshige, M. Iwata, S. Omori, Y. Tanaka, H. Hirose, K. Kaku, H. Maegawa, H. Watada, A. Kashiwagi, R. Kawamori, K. Tobe, T. Kadowaki, Y. Nakamura, S. Maeda, Association of new loci identified in European genome-wide association studies with susceptibility to type 2 diabetes in the Japanese, PLoS One 6 (2011) e26911.
[67] R. Wagner, K. Dudziak, S.A. Herzberg-Schafer, F. Machicao, N. Stefan, H. Staiger, H.U. Haring, A. Fritsche, Glucose-raising genetic variants in MADD and ADCY5 impair conversion of proinsulin to insulin, PLoS One 6 (2011) e23639.
[68] F.J. Tsai, C.F. Yang, C.C. Chen, L.M. Chuang, C.H. Lu, C.T. Chang, T.Y. Wang, R.H. Chen, C.F. Shiu, Y.M. Liu, C.C. Chang, P. Chen, C.H. Chen, C.S. Fann, Y.T. Chen, J.Y. Wu, A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese, PLoS Genet. 6 (2010) e1000847.
[69] R. Saternus, S. Pilz, S. Graber, M. Kleber, W. Marz, T. Vogt, J. Reichrath, A closer look at evolution: variants (SNPs) of genes involved in skin pigmentation, including EXOC2, TYR, TYRP1, and DCT, are associated with 25(OH)D serum concentration, Endocrinology 156 (2015) 39–47.
[70] P. Reemann, E. Reimann, S. Ilmjarv, O. Porosaar, H. Silm, V. Jaks, E. Vasar, K. Kingo, S. Koks, Melanocytes in the skin–comparative whole transcriptome analysis of main skin cell types, PLoS One 9 (2014) e115717.
[71] S.I. Candille, D.M. Absher, S. Beleza, M. Bauchet, B. McEvoy, N.A. Garrison, J.Z. Li, R.M. Myers, G.S. Barsh, H. Tang, M.D. Shriver, Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations, PLoS One 7 (2012) e48294.
[72] C. Bertolotto, Melanoma: from melanocyte to genetic alterations and clinical options, Scientifica 2013 (2013) 635203.
[73] I. Stefanaki, O.A. Panagiotou, E. Kodela, H. Gogas, K.P. Kypreou, F. Chatzinasiou, V. Nikolaou, M. Plaka, I. Kalfa, C. Antoniou, J.P. Ioannidis, E. Evangelou, A.J. Stratigos, Replication and predictive value of SNPs associated with melanoma and pigmentation traits in a Southern European case-control study, PLoS One 8 (2013) e55712.
[74] L.H. Butterfield, S.M. Jilani, N.G. Chakraborty, L.A. Bui, A. Ribas, V.B. Dissette, R. Lau, S.C. Gamradt, J.A. Glaspy, W.H. McBride, B. Mukherji, J.S. Economou, Generation of melanoma-specific cytotoxic T lymphocytes by dendritic cells transduced with a MART-1 adenovirus, J. Immunol. 161 (1998) (1950) 5607–5613 Baltimore, Md.
[75] M.F. Smikle, R. Wright-Pascoe, E.N. Barton, G. Dowe, D.T. Gilbert, E. Choo-Kang, O.S. Morgan, Autoantibodies, human T lymphotropic virus type 1 and type 1 diabetes mellitus in Jamaicans, W. Indian Med. J. 51 (2002) 153–156.
[76] C. Alves, L. Dourado, Endocrine and metabolic disorders in HTLV-1 infected patients, Braz. J. Infect. Dis. : Offic. Publ. Brazil Soc. Infect. Dis. 14 (2010) 613–620.
[77] S. Tavintharan, K. Woon, L.T. Pek, N. Jauhar, X. Dong, S.C. Lim, C.F. Sum, Niacin results in reduced monocyte adhesion in patients with type 2 diabetes mellitus, Atherosclerosis 215 (2011) 176–179.
[78] M.G. Petrovic, J. Osredkar, M. Saraga-Babic, D. Petrovic, K469E polymorphism of the intracellular adhesion molecule 1 gene is associated with proliferative diabetic retinopathy in Caucasians with type 2 diabetes, Clin. Exp. Ophthalmol. 36 (2008) 468–472.
[79] V.T. Falcao, D.A. Maschio, C.C. de Fontes, R.B. Oliveira, J.C. Santos-Silva, A.C. Almeida, E.C. Vanzela, M.T. Cartaxo, C.P. Carvalho, C.B. Collares-Buzato, Reduced insulin secretion function is associated with pancreatic islet redistribution of cell adhesion molecules (CAMS) in diabetic mice after prolonged high-fat diet, Histochem. Cell Biol. 146 (2016) 13–31.
[80] S.M. Ayuk, H. Abrahamse, N.N. Houreld, The role of photobiomodulation on gene expression of cell adhesion molecules in diabetic wounded fibroblasts in vitro, J. Photochem. Photobiol. B Biol. 161 (2016) 368–374.
[81] C.H. Tseng, Metformin and esophageal cancer risk in Taiwanese patients with type 2 diabetes mellitus, Oncotarget 8 (2017) 18802–18810.
[82] A. Spadigam, A. Dhupar, S. Syed, T.S. Saluja, Diabetes, Epstein-Barr virus and extranodal natural killer/T-cell lymphoma in India: unravelling the plausible nexus, Indian J. Med. Paediatr. Oncol. : Offic. J. Indian Soc. Med. Paediatric. Oncol. 37 (2016) 6–13.
[83] J. Smit, Community-acquired Staphylococcus aureus bacteremia: studies of risk and prognosis with special attention to diabetes mellitus and chronic heart failure, Danish Med. J. 64 (2017).
[84] C.W. Farnsworth, E.M. Schott, S.E. Jensen, J. Zukoski, A.M. Benvie, M.A. Refaai, S.L. Kates, E.M. Schwarz, M.J. Zuscik, S.R. Gill, R.A. Mooney, Adaptive upregulation of clumping factor A (ClfA) by Staphylococcus aureus in the obese, type 2 diabetic host mediates increased virulence, Infect. Immun. (2017) 85.
[85] S. Sato, M. Nishida, K. Nasu, H. Narahara, K. Norose, F. Aosai, Congenital toxoplasmosis from a mother with type 2 diabetes mellitus: a case report, J. Obstet. Gynaecol. Res. 40 (2014) 2158–2161.
[86] H. Majidiani, S. Dalvand, A. Daryani, M.L. Galvan-Ramirez, M. Foroutan-Rad, Is chronic toxoplasmosis a risk factor for diabetes mellitus? A systematic review and meta-analysis of case-control studies, Braz. J. Infect. Dis. : Offic. Publ. Brazil Soc. Infect. Dis. 20 (2016) 605–609.
[87] S. Kankova, J. Flegr, P. Calda, An elevated blood glucose level and increased incidence of gestational diabetes mellitus in pregnant women with latent toxoplasmosis, Folia Parasitol. 62 (2015).
[88] X.T. Truong, S.H. Park, Y.G. Lee, H.Y. Jeong, J.H. Moon, T.I. Jeon, Protocatechuic acid from pear inhibits melanogenesis in melanoma cells, Int. J. Mol. Sci. 18 (2017).
[89] H. Cho, O. Kim, Y. Lee, L.J. Kang, C.N. Nguyen, A. Ishihara, H.E. Kim, Feruloylserotonin inhibits hydrogen peroxide-induced melanogenesis and apoptosis in B16F10 and SK-Mel-2 melanoma cells, Biochem. Biophys. Res. Commun. 491 (2017) 973–979.
[90] M. Sivanandham, B. Mukherji, Functionally different HTLV I-infected T cell lines with the same phenotype derived from a patient with melanoma, Immunol. Lett. 23 (1989) 149–154.
[91] S. Gupta, A. Imam, Malignant melanoma in a homosexual man with HTLV-III/LAV exposure, Am. J. Med. 82 (1987) 1027–1030.
[92] K.A. Paschos, D. Canovas, N.C. Bird, The role of cell adhesion molecules in the progression of colorectal cancer and the development of liver metastasis, Cell. Signal. 21 (2009) 665–674.
[93] H. Nelson, P.S. Ramsey, J.H. Donohue, L.E. Wold, Cell adhesion molecule expression within the microvasculature of human colorectal malignancies, Clin. Immunol. Immunopathol. 72 (1994) 129–136.
[94] D.J. Rowan, C.P. Hartley, L.F. Carrillo-Polanco, K. Oshima, C.E. Hagen, Diagnostic phrasing is independently correlated with the decision to treat for graft-versus-host disease: retrospective review of colon biopsies with rare apoptosis, Histopathology 69 (2016) 802–811.
[95] A.J. Gomez, S. Arai, J.P. Higgins, N. Kambham, Clinicopathologic threshold of acute colorectal graft-versus-host disease, Arch. Pathol. Lab Med. 140 (2016) 570–577.
[96] L. Sun, S. Yu, Diabetes mellitus is an independent risk factor for colorectal cancer, Dig. Dis. Sci. 57 (2012) 1586–1597.
[97] H. Khalili, A.T. Chan, Is diabetes a risk factor for colorectal cancer? Dig. Dis. Sci. 57 (2012) 1427–1429.
[98] L. Deng, Z. Gui, L. Zhao, J. Wang, L. Shen, Diabetes mellitus and the incidence of colorectal cancer: an updated systematic review and meta-analysis, Dig. Dis. Sci. 57 (2012) 1576–1585.

[99] E. Jullumstro, M. Kollind, S. Lydersen, T.H. Edna, Diabetes mellitus and outcomes of colorectal cancer, Acta Oncol. (Stockholm, Sweden) 48 (2009) 361–367.
[100] T. Yamamoto, S. Tanaka, Y. Nakamura, R. Hashiba, K. Ishihara, S. Kubo, K. Hirohashi, K. Ohno, Colorectal cancer metastasis to the thyroid, Osaka City Med. J. 58 (2012) 67–75.
[101] B. Boursi, K. Haynes, R. Mamtani, Y.X. Yang, Thyroid dysfunction, thyroid hormone replacement and colorectal cancer risk, J. Natl. Cancer Inst. 107 (2015) djv084.
[102] B. Li, A.P. Castano, T.E. Hudson, B.T. Nowlin, S.-L. Lin, J.V. Bonventre, K.D. Swanson, J.S. Duffield, The melanoma-associated transmembrane glycoprotein Gpnmb controls trafficking of cellular debris for degradation and is essential for tissue repair, FASEB J. 24 (2010) 4767–4781.
[103] S. Calderari, J.C. Irminger, M.H. Giroix, J.A. Ehses, M.N. Gangnerau, J. Coulaud, K. Rickenbach, D. Gauguier, P. Halban, P. Serradas, F. Homo-Delarche, Regenerating 1 and 3b gene expression in the pancreas of type 2 diabetic Goto-Kakizaki (GK) rats, PLoS One 9 (2014) e90045.
[104] M. Lohr, B. Bergstrome, R. Maekawa, M.B. Oldstone, G. Kloppel, Human cytomegalovirus in the pancreas of patients with type 2 diabetes: is there a relation to clinical features, mRNA and protein expression of insulin, somatostatin, and MHC class II? Virchows Archiv - A Pathol. Anat. Histopathol. 421 (1992) 371–378.
[105] H. Ruihua, Z. Mengyi, Z. Chong, Q. Meng, M. Xin, T. Qiulin, B. Feng, L. Ming, RhoA regulates resistance to irinotecan by regulating membrane transporter and apoptosis signaling in colorectal cancer, Oncotarget 7 (2016) 87136–87146.
[106] Y. Chen, B.S. Harrington, K.C. Lau, L.J. Burke, Y. He, M. Iconomou, J.S. Palmer, B. Meade, J.W. Lumley, J.D. Hooper, Development of an enzyme-linked immunosorbent assay for detection of CDCP1 shed from the cell surface and present in colorectal cancer serum specimens, J. Pharm. Biomed. Anal. 139 (2017) 65–72.
[107] K. Nishi, H. Luo, K. Nakabayashi, K. Doi, S. Ishikura, Y. Iwaihara, Y. Yoshida, K. Tanisawa, T. Arai, S. Mori, M. Sawabe, M. Muramatsu, M. Tanaka, T. Sakata, S. Shirasawa, T. Tsunoda, An alpha-kinase 2 gene variant disrupts filamentous actin localization in the surface cells of colorectal cancer spheroids, Anticancer Res. 37 (2017) 3855–3862.
[108] S. Tiwari, B. Tirosh, A. Rubinstein, Increasing the affinity of cationized polyacrylamide-paclitaxel nanoparticles towards colon cancer cells by a surface recognition peptide, Int. J. Pharm. 531 (2017) 281–291.
[109] S. Cortes, S. Chambers, A. Jeronimo, D. Isenberg, Diabetes mellitus complicating systemic lupus erythematosus - analysis of the UCL lupus cohort and review of the literature, Lupus 17 (2008) 977–980.
[110] G. Orozco, P. Eerligh, E. Sanchez, S. Zhernakova, B.O. Roep, M.A. Gonzalez-Gay, M.A. Lopez-Nevot, J.L. Callejas, C. Hidalgo, D. Pascual-Salcedo, A. Balsa, M.F. Gonzalez-Escribano, B.P. Koeleman, J. Martin, Analysis of a functional BTNL2 polymorphism in type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus, Hum. Immunol. 66 (2005) 1235–1241.
[111] J.F. Ludvigsson, A. Rubio-Tapia, V. Chowdhary, J.A. Murray, J.F. Simard, Increased risk of systemic lupus erythematosus in 29,000 patients with biopsy-verified celiac disease, J. Rheumatol. 39 (2012) 1964–1970.
[112] S.W. Park, S.J. Park, J.I. Shin, Another possible underlying mechanism for the positive association between celiac disease and systemic lupus erythematosus: the role of interleukin 21, J. Rheumatol. 40 (2013) 1619.
[113] K. Ben Abdelghani, L. Mouelhi, A. Hriz, S. Hajri, T. Najjar, M. Mahfoudhi, S. Turki, A. Khedher, Systemic lupus erythematosus and celiac disease, Joint, bone, spine : revue du rhumatisme 79 (2012) 202–203.
[114] H.J. Freeman, Adult celiac disease followed by onset of systemic lupus erythematosus, J. Clin. Gastroenterol. 42 (2008) 252–255.
[115] T. Ishikawa, M. Koshizaka, Y. Maezawa, M. Takemoto, Y. Tokuyama, T. Saito, K. Yokote, Continuous glucose monitoring reveals hypoglycemia risk in elderly patients with type 2 diabetes mellitus, J. Diabet. Invest. (2017) 69–74.
[116] V. Shriraam, S. Mahadevan, M. Anitharani, N.S. Jagadeesh, S.B. Kurup, T.A. Vidya, K.G. Seshadri, Reported hypoglycemia in Type 2 diabetes mellitus patients: prevalence and practices-a hospital-based study, Indian J. Endocrinol. Metabol. 21 (2017) 148–153.
[117] S. Kachroo, H. Kawabata, S. Colilla, L. Shi, Y. Zhao, J. Mukherjee, U. Iloeje, V. Fonseca, Association between hypoglycemia and fall-related events in type 2 diabetes mellitus: analysis of a U.S. commercial database, J. Manag. Care Spec. Pharm. 21 (2015) 243–253.
[118] H. Dai, W.Q. Li, A.A. Qureshi, J. Han, Personal history of psoriasis and risk of nonmelanoma skin cancer (NMSC) among women in the United States: a population-based cohort study, J. Am. Acad. Dermatol. 75 (2016) 731–735.
[119] A. Egeberg, J.P. Thyssen, G.H. Gislason, L. Skov, Skin cancer in patients with psoriasis, J. Eur. Acad. Dermatol. Venereol. : JEADV 30 (2016) 1349–1353.
[120] M.S. Lee, R.Y. Lin, Y.T. Chang, M.S. Lai, The risk of developing non-melanoma skin cancer, lymphoma and melanoma in patients with psoriasis in Taiwan: a 10-year, population-based cohort study, Int. J. Dermatol. 51 (2012) 1454–1460.
[121] M. Hashemzadeh, M.R. Movahed, The occurrence of asthma in hospitalized patients with type 2 diabetes mellitus, Intern. Med. J. 39 (2009) 699–701.
[122] A.E. Mirrakhimov, S.G. Kwatra, T.S. Shah, Should we consider asthma as a risk factor for diabetes mellitus? Mayo Clin. Proc. 88 (2013) 211.
[123] P. Ungprasert, K. Wijarnpreecha, W. Kittanamongkolchai, Psoriasis and risk of celiac disease: a systematic review and meta-analysis, Indian J. Dermatol. 62 (2017) 41–46.
[124] L. Abenavoli, L. Leggio, G. Gasbarrini, G. Addolorato, Celiac disease and skin: psoriasis association, World J. Gastroenterol. 13 (2007) 2138–2139.
[125] V. Ojetti, J. Aguilar Sanchez, C. Guerriero, B. Fossati, R. Capizzi, C. De Simone, A. Migneco, P. Amerio, G. Gasbarrini, A. Gasbarrini, High prevalence of celiac disease in psoriasis, Am. J. Gastroenterol. 98 (2003) 2574–2575.
[126] M. Vidal, M.E. Cusick, A.L. Barabasi, Interactome networks and human disease, Cell 144 (2011) 986–998.
[127] K.I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, A.L. Barabasi, The human disease network, Proc. Natl. Acad. Sci. U. S. A 104 (2007) 8685–8690.
[128] A.L. Barabasi, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nat. Rev. Genet. 12 (2011) 56–68.
[129] X. Wu, R. Jiang, M.Q. Zhang, S. Li, Network-based global inference of human disease genes, Mol. Syst. Biol. 4 (2008) 189.
[130] F. Barrenas, S. Chavali, P. Holme, R. Mobini, M. Benson, Network properties of complex human disease genes identified through genome-wide association studies, PLoS One 4 (2009).