Journal of Autoimmunity 97 (2019) 48–58

Contents lists available at ScienceDirect

Journal of Autoimmunity

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

Identification of novel susceptibility genes associated with seven autoimmune disorders using whole genome molecular interaction networks

Sam Kara[a,c], Gerardo A. Pirela-Morillo[b], Conrad T. Gilliam[a], George D. Wilson[c,*]

[a] University of Chicago, Departments of Human Genetics, 920 East 58 th St., Chicago, IL 60637, USA
[b] La Universidad del Zulia, Computer Science Department, Laboratories for Computational Models & Languages, and Bioinformatics, Edif. Grano de Oro, Planta Baja, Departamento de Computación, Ave. Universidad con Ave. 22, Maracaibo, 4002, Venezuela
[c] Radiation Oncology Department, Beaumont Health, 3811 W Thirteen Mile Road, Royal Oak, MI, 48073, USA

ARTICLE INFO

Keywords:
Autoimmune disorders
Crohn's disease
Molecular network
Psoriasis
Type-1 diabetes
Rheumatoid arthritis

ABSTRACT

Convergent evidence from multiple and independent genetics studies implicate a small number of genes that predispose individuals to multiple autoimmune disorders (AuD). These intersecting loci reinforced the hypothesis that disorders with overlapping etiology group into a cluster of closely related genes within a whole genome molecular interaction network. We tested the hypothesis that "biological network proximity" within a whole genome molecular interaction network can be used to inform the search for multigene inheritance. Using a set of nine previously published genome wide association studies (GWAS) of AuD genes, we generated AuD-specific molecular interaction networks to identify networks of associated genes. We show that all nine "seed genes" can be connected within a 35-member network via interactions with 26 connecting genes. We show that this network is more connected than expected by chance, and 13 of the connecting genes showed association with multiple AuD upon GWAS reanalysis. Furthermore, we report association of SNPs in five new genes (IL10RA, DGKA, GRB2, STAT5A, and NFATC2) which were not previously considered as AuD candidates, and show significant association in novel disease samples of Crohn's disease and systemic lupus erythematosus. Furthermore, we show that the connecting genes show no association in four non-AuD GWAS. Finally, we test the connecting genes in psoriasis GWAS, and show association to previously identified loci and report new loci. These findings support the hypothesis that molecular interaction networks can be used to inform the search for multigene disease etiology, especially for disorders with overlapping etiology.

1. Introduction

Genome-wide association studies (GWAS) have revolutionized our approach for mapping and identifying genetic factors associated with common and complex human disorders. The advances made in mapping DNA loci for diseases simultaneously with the mapping of DNA loci for disease trait has uncovered many genetics loci that associate with human diseases [1–4]. However, the disease risk imparted by individual genotypes tends to be very small and consequently of limited predictive value. Understanding the relationships between multigene genotypes and disease risk is therefore essential for elucidating the molecular networks and biological process underlying complex diseases [5–9]. Our group and others have predicted that genes predisposing to more than one clinical disorder will map more proximal to one another within a whole genome molecular interaction network. This "biological network proximity" can consequently be used in combination with GWAS data to explore the genetic architecture of common and complex human disorders [5,8,10–12]. In order to test this hypothesis, we targeted six commonly known heritable autoimmune diseases (AuD) characterized by high convergence of genotype-phenotype predictions across multiple independent GWAS.

The cause(s) of the commonly known AuD remain largely unknown. It has long been recognized that environmental influences and genetic factors play an important role in disease risk [13,14]. Interestingly, AuD

Abbreviations: AuD, autoimmune disorders; GWAS, genome wide association studies; SNPs, single nucleotide polymorphisms; CD, Crohn's disease; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1D, type-1 diabetes; T2D, type-2 diabetes; CeD, celiac disease; MS, multiple sclerosis; TRIDOM, Translational Research Initiative in the Department of Medicine at the University of Chicago; IPA, Ingenuity Pathway Analysis; WTCCC, the Wellcome Trust Case-Control Consortium; NIDDK, the National Institute of Digestive Disorders and Kidney Disease; IMSGC, the International Multiple Sclerosis Genomics Consortium; SLEGEN, the Whole Genome Association Study of Systemic Lupus Erythematosus; PSO, Psoriasis; LD, linkage disequilibrium

* Corresponding author. William Beaumont Hospital, Department of Radiation Oncology, 3811 W Thirteen Mile Road, Royal Oak, MI, 48073, USA.
E-mail address: george.wilson@beaumont.edu (G.D. Wilson).

https://doi.org/10.1016/j.jaut.2018.10.002
Received 4 September 2018; Received in revised form 8 October 2018; Accepted 11 October 2018
Available online 01 November 2018
0896-8411/

have significantly higher inherited predisposition and can co-occur within families and even within individual patients, bolstering the hypothesis that clinically distinct AuD may share common genetic risk loci [11,15–18]. In support of this, genetic associations have been observed in PTPN2 and PTPN22 genes encoding protein tyrosine phosphatase, non-receptor type 2 and 22, respectively, with Crohn's disease (CD) [19], rheumatoid arthritis (RA) [20,21], systemic lupus erythematosus (SLE) [22], and type-1 diabetes (T1D) [23]; IL2 and IL21 genes encoding interleukin2 and interleukin 21, respectively, with celiac disease (CeD) [24,25], multiple sclerosis (MS) [26], T1D [23], and RA [20,27]; IFIH1 gene encoding the interferon induced with helicase C domain 1 with SLE [28] and T1D [29,30]; IL12A gene encoding interleukin 12A with MS, T1D [31], and CeD [31]; and SH2B3 gene encoding SH2B adaptor protein 3 with CeD [32] and T1D [33]. In addition, SH2B3 gene encoding SH2B adaptor protein 3 is associated with CeD [32] and T1D [33]. AuD thus provide an ideal scenario to evaluate whether genome-wide molecular interaction networks can be used to inform the search for multigenic determinants of common heritable disorders.

A number of recent studies have described the rationale for using 'molecular networks' or 'molecular systems' to explore the genetic architecture of complex, multigenic biological processes including common heritable disorders in humans [10,34,35]. Other studies suggest that metrics such as 'network proximity' that measure the number of molecular interactions separating individual molecules within a molecular interaction network might be used to inform the search for multigene etiologies of disease [8,10]. Moreover, public availability of GWAS data using various technology platforms for a wide range of diseases in well-characterized and non-overlapping clinical samples, provide an opportunity to quickly evaluate genetic evidence for novel candidate disease genes that map proximal to known genes within a molecular interaction network.

Here, we generate AuD-specific molecular interaction networks using nine previously reported AuD candidate genes that have been implicated in at least two of six commonly known AuDs, namely CD, CeD, MS, T1D, SLE, and RA [11]. We hypothesized that this commonality would be reflected by a non-random, proximal molecular interaction relationship among these nine AuD "seed genes" and further, that the minimal set of "network connecting genes" required to connect the maximal number of seed genes would be enriched as targets for AuD-predisposing genetic risk variants. We demonstrate that molecular interaction properties such as network proximity can serve as proxies for the type of multigene patterns of inheritance that give rise to common, heritable disorders in humans.

2. Methods

2.1. Study design

We use "network proximity" to identify a small number of candidate genes that we then "re-evaluated" in the published GWAS studies. By lowering the number of single nucleotide polymorphisms (SNPs) tested we sought to detect candidate AuD genes that were indistinguishable from background noise in the genome wide studies. We used this strategy to winnow a set of 26 candidate genes to a small number that could be evaluated in a novel case: control study. Thus, our study design is based on the identification and association analysis of a very small number of candidate genes (relative to a whole genome scan) where the statistical cost of multiple testing is greatly reduced.

2.2. Selection of candidate "seed" genes and AuD dataset

Our goal was to select a small number of previously validated and well replicated AuD genes to seed the formation of a molecular interaction gene network that we hypothesized would be enriched for other AuD candidate genes. We selected nine previously published GWAS-AuD genes drawn from 18 GWAS and are common among two or more of the following six AuD: CeD, CD, MS, RA, SLE and T1D [11]. These nine "seed genes" represent a subset of the hundreds of identified and replicated genes from various studies and does not represent a summary of all AuD gene interactions. The GWAS for the six AuD was requested from multiple sources. Approved access from GWAS datasets were used (See section 3.3).

2.3. Gene identification

A gene location was defined to include 100-kilobase up-stream and down-stream of NCBI's start- and end-gene location (Version 10.2). All examined SNPs and their physical coordinates were downloaded from each database and where assigned to the prospective gene(s) as defined above to enable SNP-gene comparisons across multiple databases and platforms.

2.4. Graphical representation of gene networks

The term 'network' is used to refer to a graphical representation of the molecular relationships between genes or gene products. Genes or gene products are represented as nodes (shapes) and the biological relationship between two nodes is represented as an edge (line). All nodes and edges are supported by at least one reference from the literature, from a textbook, or from a database that was incorporated into the Ingenuity knowledge base. Nodes are displayed using various shapes that represent the experimentally determined or gene ontology inferred functional class of the gene product. In order to facilitate visualization of the seed and network connecting genes we only show the molecular interactions (edges) connecting network members. As described in the text, we tasked the IPA software to "link together the maximum number of seed genes with a minimal number of connecting genes within the constraints of the default 35-node network". It is expected that optimization of this problem will include one or more nodes with the network property of a "hub"; i.e., the gene may be selected based on its connections to a large number of molecules rather than biological similarity to the other network genes. Ingenuity classifies molecular interactions as either "direct" or "indirect". Direct interactions refer to actions like "binding", "cleave", or "phosphorylate" whereas indirect interactions refer to actions like "activate", "inhibit", or "stimulate". In this study, we only consider direct interaction. We used Pathway Studio 9 MammalPlus (Elsevier B.V.) program for network graphical representation and to reduce the number of edges for clarity of the graphical representation.

2.5. Molecular network analysis

We used the Ingenuity Pathway Analysis (IPA) software (release 7.0) (Ingenuity Systems, Mountain View, CA, USA) and Knowledge Base to predict molecular interaction relationships among the nine AuD seed genes and to predict the connecting genes. Seed genes were selected as described above and the corresponding refseq accession numbers were uploaded to the IPA program for analysis. The IPA algorithm first searches for evidence of direct interaction between seed genes until the maximum number of seed genes are incorporated into the default 35-member network. The IPA software calculates the probability that the resulting network occurred by chance; i.e., that the final network is comprised of a random collection of genes. A score of 2 indicates there is a 1/100 chance ($p < 0.05$; 99% confidence level) that the listed group of genes were incorporated randomly into the molecular interaction network.

The IPA software uses Gene Ontology classifications to compare and rank-order gene function predictions for all genes within a given network. The top-ranking predictions for the 35-member gene network are listed in Table 1. The software also calculates whether, and to what extent, gene function classifications among the network members differs from the full collection of genes in the Ingenuity pathway

Table 1
The seed and network genes, their chromosomal location, and function.

Gene symbol	Chr.	Function
Seed genes		
IFIH1	2	Interferon induced with helicase C domain 1
IL12A	3	Interleukin 12A
IL2	4	Interleukin 2
IL21	4	Interleukin 21
IL2RA	10	Interleukin 2 receptor subunit alpha
PTPN2	18	Protein tyrosine phosphatase, non-receptor type 2
PTPN22	1	Protein tyrosine phosphatase, non-receptor type 22
SH2B3	12	SH2B adaptor protein 3
STAT4	2	Signal transducer and activator of transcription 4
Network genes		
ATG12	5	Autophagy related 12
CXCL9	4	C-X-C motif chemokine ligand 9
DGKA	12	Diacylglycerol kinase alpha
EGFR	7	Epidermal growth factor receptor
ERRFI1	1	ERBB receptor feedback inhibitor 1
GRB2	17	Growth factor receptor bound protein 2
IL10RA	11	Interleukin 10 receptor subunit alpha
IL12B	5	Interleukin 12B
IL12RB1	19	Interleukin 12 receptor subunit beta 1
IL12RB2	1	Interleukin 12 receptor subunit beta 2
IL17A	6	Interleukin 17A
IL18R1	2	Interleukin 18 receptor subunit 1
IL18RAP	2	Interleukin 18 receptor accessory protein
IL23A	12	Interleukin 23A
LRRFIP1	2	LRR binding FLII interacting protein 1
NFATC2	20	Nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 2
NFKB1	4	Nuclear factor kappa B subunit 1
NFKB2	10	Nuclear factor kappa B subunit 2
REL	2	REL proto-oncogene, NF-kB subunit
RELA	11	RELA proto-oncogene, NF-kB subunit
RELB	19	RELB proto-oncogene, NF-kB subunit
RORC	1	RAR-related orphan receptor gamma
SOCS7	17	Suppressor of cytokine signaling 7
STAT3	17	Signal transducer and activator of transcription 3
STAT5A	17	Signal transducer and activator of transcription 5A
ZNF467	7	Zinc finger protein 467

knowledge base reference set. The program uses a right tailed Fisher's exact test (with a $p < 0.05$) to calculate the probability that each biological function assigned to two or more network genes occurred by chance.

2.6. The University of Chicago TRIDOM-Crohn's disease (TRIDOM-CD) patients and samples

Genomic DNA samples were obtained from the Translational Research Initiative in the Department of Medicine (TRIDOM) at the University of Chicago. A total of 376 CD patients (176 female, 198 male) were of European-American (n = 336), black and African-American (n = 19), Hispanic (n = 4), Asian (n = 3), and Indian-American (n = 1) ancestries. All patients met the American Gastroenterological Association criteria for diagnosis of CD and SLE. The study was approved by the institutional review boards at all institutions, and informed consent was obtained from all subjects in the study. For this particular study, only European-American patients were considered. A total of 384 DNA samples of European-American ancestry used previously as control individuals for the study of anxiety and related disorders were obtained from Columbia University and used as controls for the present study [36]. The observed genotype frequency in cases and controls did not deviate from Hardy-Weinberg proportions ($p > 0.01$).

2.7. The University of Chicago TRIDOM-Human systemic lupus erythematosus (TRIDOM-SLE)

Genomic DNA samples were obtained from the TRIDOM and Rush University Medical Center. Of the 492 SLE patients, 236 were African American, 136 were European American, and 94 were Hispanic American. African American controls (n = 140) from the TRIDOM registry were also genotyped, and these subjects were screened by medical record review for the absence of autoimmune or inflammatory disease by the same physician (LR) [37]. The study was approved by the institutional review board at each institution, and informed consent was obtained from all subjects. The observed genotype frequency in cases and controls did not deviate from Hardy-Weinberg proportions ($p > 0.01$).

2.8. SNP genotyping

The CD tested SNPs were genotyped using the Sequenom iPLEX™ assay and MassArray platform according to the manufacturer's protocols (Sequenom, San Diego, CA). SNPs genotyped in the SLE sample were genotyped using the ABI7900 according to the manufacturer's protocols.

2.9. Statistical analysis

The genotype-phenotype genetic association p-values were extracted from the GWAS databases described above. We estimated the number of independent SNPs for each gene, using the method of Nyholt [38] which performs spectral decomposition of matrices of pairwise LD between SNPs. HapMap and 1000-genome data were used to estimate LD. The AuD datasets studies used by us, a statistical significance was defined by $p < 0.05$. We applied a Bonferroni correction, based on the number of independent SNPs, to all p-values less than 0.05 and reported both the smallest uncorrected p-value for each gene (extracted from the original study) along with the smallest corrected value (extracted from the original study), the total number of valid SNPs genotyped, and the total number of SNPs per gene with corrected p-values less than 0.05. To evaluate evidence of genotype-phenotype association, we selected the smallest adjusted p-value (reported in each study) and applied a Bonferroni correction based on the number of independent SNPs. The association between the typed SNPs and both CD and SLE in the TRIDOM sample was investigated using a chi-square allelic test.

3. Results

3.1. Whole genome molecular network identifies 26 candidate autoimmune disorder genes

The nine AuD “seed genes” were selected based on the results of 18 published GWAS published in 2008 or prior [11]. Each of the nine AuD seed genes as associated with two or more of the following six AuD: CeD, CD, MS, RA, SLE and T1D [11]. The nine seed genes selected included: PTPN22 (CD, T1D, RA, and SLE), IFIH1 (T1D and RA), STAT4 (SLE and RA), IL12A (CeD and T1D), Il2 (CeD, T1D, RA and MS), Il21 (CeD, T1D, RA and MS), IL2RA (T1D and MS), SH2B3 (T1D and CeD) and PTPN2 (T1D and CD). IPA software was used to examine the connectivity of these nine seed genes among each other which resulted in a single network connecting all nine seed genes. The software was used to build a 35-member network containing the maximum number of directly connected AuD seed genes (Table 1). Fig. 1 depicts the network from this search that shows a 35-member network comprised of nine seed genes and 26 “network connecting genes”. For illustration purposes, only few connections among genes are shown. Supplement table 1 list all interactions among the 35-member network. The Ingenuity database provided information regarding a given gene's likely cellular location and allows classification of genes and groups of genes by standard gene ontology classifications. The fact that a number of the candidate seed genes and network connecting genes are designated as G-protein coupled receptors or growth factors with established roles in signaling, cellular development, cell growth and proliferation, and

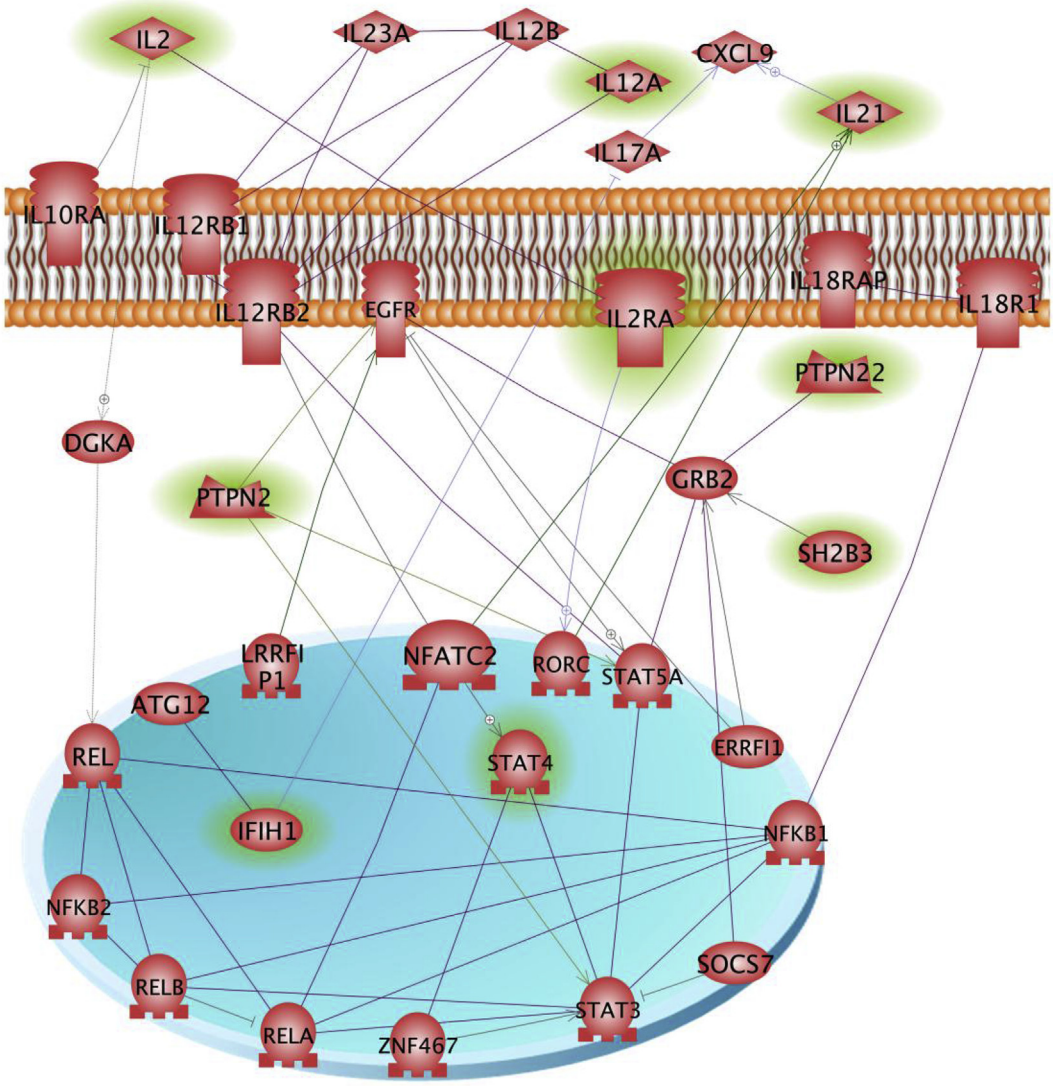

Fig. 1. Autoimmune disease specific molecular interaction network. Seed genes (highlighted in gray) and candidate genes are displayed in their identified cellular compartment for six autoimmune diseases (CeD, CD, MS, RA, SLE and T1D). Genes or gene products are represented as nodes/shapes, and the biological relationship between two nodes is represented as an edge (line). Genes highlighted in green represent the seed genes. All nodes and edges are supported by at least 1 reference from the literature, from a textbook, or from a database that was incorporated into Ingenuity knowledge base. Nodes are displayed using various shapes that represent the functional class of the gene product or molecule class. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

inflammatory response provides evidence of biological relevance.

3.2. The AuD network did not arise by chance

IPA calculated the probability that the AuD network could arise by chance by using nine "randomly matched" seed genes. When compared to random permutation, the 35-member AuD network was found to be statistically significant with a score of 26 ($p < 10^{-26}$), where a score of 2 indicates there is a 1/100 chance that the observed network would occur by chance ($p < 0.05$; 99% confidence level). Furthermore, we estimated the statistical likelihood that randomly selected groups of nine genes would produce a similar result by performing 100 independent simulations. Using the identical software parameters as those used to generate the network depicted in Fig. 1, we generated random sets of 9-genes matched for number of edges and connections from the 20,000 plus genes in the NCBI gene database. We then "seeded" 100 independent simulations with the 9-gene sets and tasked the IPA software to identify the 35-member network with the maximum number of seed genes. The simulation results ranged from 0 to 4 seed genes per network (Fig. 2). The IPA user privileges prohibit our performing sufficient independent simulations to generate statistically significant confidence intervals; however, the limited number of simulations we did obtain indicates that the incorporation of the nine AuD seed genes into a single 35-member network is unlikely to have occurred by chance.

Gene Ontology comparisons of the 35-network members with all genes in the IPA interaction database suggest that the seed genes within the network showed a higher connectivity than expected by chance. For example, comparison of the network connecting genes revealed that all connecting genes have well-established roles in cell-to-cell signaling ($p < 5.28 \times 10^{-17}$) and interactions ($p < 1.0 \times 10^{-15}$), cellular development ($p < 2.44 \times 10^{-14}$) and immunological diseases ($p < 3.19 \times 10^{-13}$). These results suggest highly interacting 26 loci with the nine seed genes beyond what is expected by chance, and the common cellular location and biological function and that common risk variants encoded by members of highly connected networks might

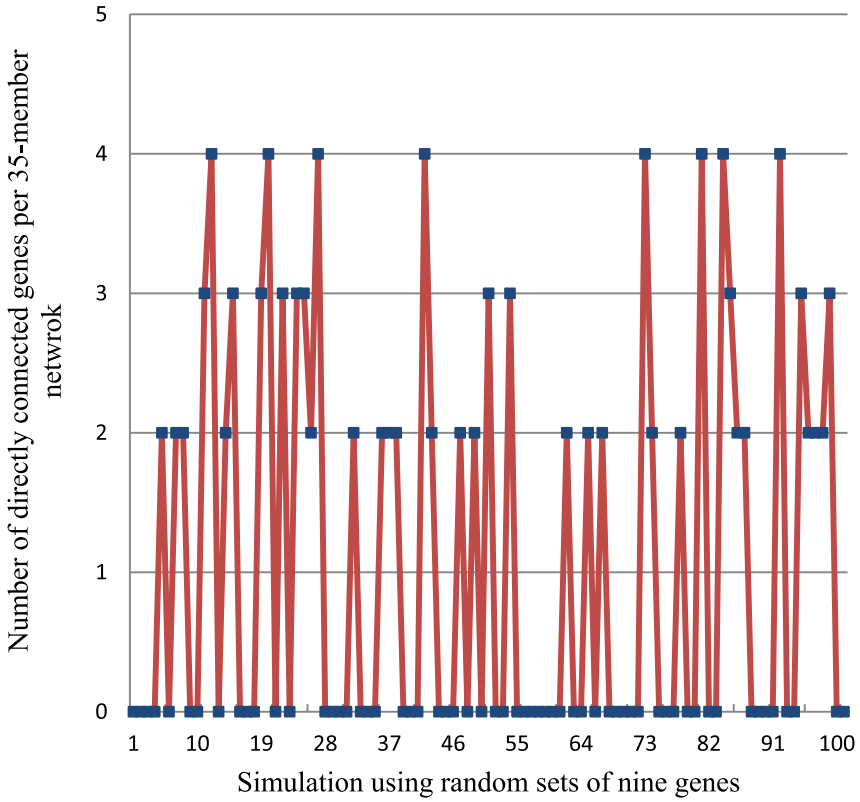

Fig. 2. Hundred random simulations of nine randomly selected genes.

possibly impact the function of a few connected genes in the same network and predispose to similar disease etiology or similar disease process.

3.3. Analysis of network connecting genes in published genome-wide association study data sets identifies 16 novel associations

To test the hypothesis that one or more of the 26 connecting network genes identified in our AuD-specific molecular interaction network is associated with AuD, we interrogated the following GWAS data resources: the Wellcome Trust Case-Control Consortium (WTCCC) and the National Institute of Digestive Disorders and Kidney Disease (NIDDK) (for CD, RA, and T1D); the Celiac Disease study (for CeD); the International Multiple Sclerosis Genomics Consortium (IMSGC) (for MS); and the Whole Genome Association Study of Systemic Lupus Erythematosus (SLEGEN) (for SLE). Fig. 3 and Supplemental Tables 1–7 list the nine seed genes and 26 connecting network genes, and association results for the previously published and publicly available GWAS. We tallied all SNPs within, and immediately surrounding, each of the candidate genes and adjusted the genotype-phenotype correlation scores for multiple testing using a conservative Bonferroni correction (see Materials and Methods). Fig. 3 summarizes the data for the presence of the significant SNPs in each disease whilst Supplemental Tables 1–7 also lists the chromosomal location of each gene, the SNP ID corresponding to the smallest *p*-value detected per gene in each study; the smallest adjusted *p*-values for the highest-ranking SNPs; and the total number of SNPs genotyped in each study; SNPs with adjusted *p*-values less than 0.05 are highlighted in bold.

3.4. Five novel genes associated with Crohn's disease

As shown in Supplemental Table 1, three of the nine seed genes show evidence of association to CD, including PTPN2 in the WTCCC data set ($p = 1.32 \times 10^{-6}$) and IL2 and IL21 in the NIDDK-IBD data set ($p = 0.031$ and $p = 0.028$, respectively). PTPN2 was previously associated with CD whereas IL2 and IL21 were not associated at the time of this study. Furthermore, we detected evidence for association ($p < 0.05$) in five (IL12RB2, IL12B, STAT3, STAT5A, and NFATC2) of the 26 network connecting genes in the WTCCC-CD data set after correcting for multiple testing (Table 1). Similarly, in the NIDDK-CD data set, two (IL12RB2 and EGFR) of the 26 connecting genes showed evidence of significant association (Table 1). Of the six network connecting genes showing evidence for association to CD in the WTCCC-CD and/or NIDDK-CD databases, three (EGFR, STAT5A and NFATC2) have not been previously implicated in CD and therefore represent entirely novel genes. One gene (IL12RB2) was not implicated in CD at the time of our study but was implicated subsequent to our analysis, which validates our approach. Two genes (IL12B and STAT3) were previously implicated and our approach confirms these previous findings.

3.5. Replication analysis of candidate genes in an independent Crohn's disease cohort

Our characterization of an AuD-specific molecular network identified novel CD candidate genes in two independent studies; the WTCCC-CD and NIDDK-CD (Supplemental Table 2). To further evaluate these findings, we used a novel CD cohort and genotyped a subset of the most significant SNPs identified in the WTCCC-CD database. The loci examined include the EGFR (rs12538371); IL12B SNP (rs1363670), STAT5A SNP (rs3816769), IL12RB2 SNPs (rs12119179 and rs56194793), IL10RA SNP (rs947889), and NFATC2 SNP (rs880324). Table 3 lists the candidate genes, along with their chromosomal locations, tested SNP, and association results for the TRIDOM-CD cohort study. We note that among those SNPs depicted in the candidate network connecting genes (Supplemental Table 1), there was no direct evidence implicating a single SNP in both of the public data sets. In light of the small size of our replication sample, we therefore followed up with only the significantly associated SNPs in the replication study

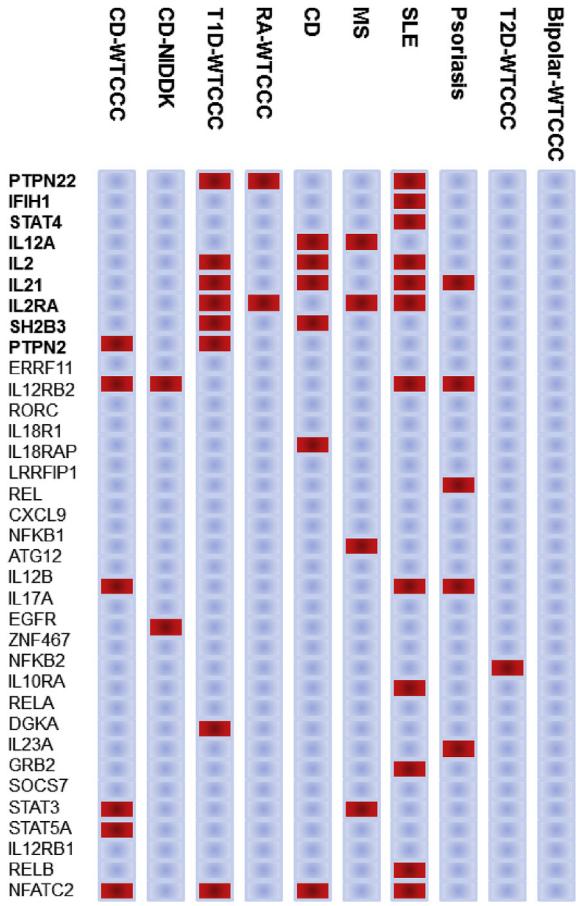

Fig. 3. The distribution of the most significant SNPs associated with each disease.

Table 2
Replication association analysis of significant SNPs in the WTCCC-CD and NIDDK-CD data sets in a novel cohort of 336 TRIDOM-Crohn's disease (TRIDOM-CD) cases and 384 control subjects. The individual SNPs showing strongest evidence for allelic association with Crohn's disease in the WTCCC-CD and NIDDK data sets were genotyped in a novel sample of 336 CD cases and 384 control subjects. Highlighted SNP IDs reach significant $p < 0.05$.

SNP ID	Gene		Allele counts			Sample counts	*p*-value
rs12119179	IL12RB2		CC	AC	AA		**0.037**
		Controls	23	163	176	362	
		Cases	40	169	158	367	
rs1363670	IL12B		GG	GC	CC		1.0
		Controls	8	81	259	348	
		Cases	7	77	243	327	
rs12538371	EGFR		AA	AG	GG		**0.021**
		Controls	221	124	17	362	
		Cases	185	136	26	347	
rs744166	STAT3		AA	AG	GG		**0.0005**
		Controls	111	169	82	362	
		Cases	132	151	45	328	
rs3816769	STAT5A		GG	AG	AA		**0.028**
		Controls	36	159	168	363	
		Cases	50	160	138	348	
rs880324	NFATC2		GG	AG	AA		**0.016**
		Controls	218	123	22	363	
		Cases	222	97	10	329	

(Table 2). The minor alleles of rs12119179 and rs56194793 in the IL12RB2 showed association with CD (p = 0.037 and p = 0.031, respectively) in the TRIDOM cohort (Table 2). This finding is consistent with association to CD ($p = 0.1.35 \times 10^{-9}$) in the WTCCC-CD data set,

Table 3
Replication association analysis of significant SNPs in the SLE data set in a novel cohort of 336 TRIDOM-Crohn's disease (TRIDOM-CD) cases and 384 control subjects. The individual SNPs showing strongest evidence for allelic association with SLE data sets were genotyped in a novel sample of 336 CD cases and 384 control subjects. Highlighted SNP IDs reach significant $p < 0.05$.

SNP ID	Gene		Allele counts			Sample counts	*p*-value
rs56194793	IL12RB2		TT	TC	CC		**0.031**
		Controls	0	69	245	314	
		Cases	0	95	220	315	
rs204471	RELB		CC	CT	TT		**0.003**
		Controls	317	45	1	363	
		Cases	308	20	0	328	
rs947889	IL10RA		GG	AG	AA		0.70
		Controls	64	176	122	362	
		Cases	67	148	113	328	
rs16967789	GRB2		AA	AG	GG		0.780
		Controls	9	97	256	362	
		Cases	9	83	237	329	

which corroborates our findings in the TRIDOM sample. The rs744166 and rs880324 SNPs in the novel connecting candidate genes (STAT5A and NFATC2 respectively) reveled significant association in the TRIDOM-CD cohort (Table 2). The association in the candidate genes further validates our prediction that STAT5A, and NFATC2 genes, or possibly a nearby gene(s), are a CD-associated gene (Supplemental Table 6). Likewise, the IL12RB2 SNPs generate a significant p-value and add to the evidence that the previously reported gene (IL12RB2) and our novel genes are implicated in CD etiology.

3.6. Replication analysis of SLE candidate genes in an independent Crohn's disease cohort

Our characterization of an AuD-specific molecular network identified novel and previously identified SLE candidate genes in (Supplemental Table 4). To test the association of these SLE candidate genes in a different AuD, in particular CD, we used a novel CD cohort and genotyped a subset of the most significant SNPs identified in the SLE database. The loci examined include the IL12RB2 (rs56194793), RELB (rs204471), IL10RA (rs947889), and GRB2 (rs16967789). Table 3 list the genotyped SNPs, candidate genes, and association results for the TRIDOM-CD cohort study. Again, we examined only the significantly associated SNPs from Supplemental Table 4 in the replication study. The minor alleles of rs56194793 in the IL12RB2 and rs204471 in the RELB showed association with CD (p = 0.031 and p = 0.003, respectively) in the TRIDOM cohort (Table 3). The tested IL10RA and GRB2 SNPs revealed no association to CD in the TRIDOM-CD cohort study.

3.7. Type-1 diabetes genome-wide association study data set

T1D has been implicated with six of the nine seed genes (PTPN22, IFIH1, IL12A, IL2, IL21, IL2RA, SH2B3, and PTPN2) at the start of our study. Supplemental Table 2 lists the seed genes along with their chromosomal location, the smallest p-value SNP (SNP ID) along with its adjusted p-value and the total number of SNPs per gene. We detected evidence for association ($p < 0.05$) to T1D in two (DGKA and NFATC2) of the 26 network connecting genes in the WTCCC-T1D data set, after correcting for multiple testing within each gene. Both DGKA and NFATC2 have not been previously implicated in CD and therefore represent entirely novel genes.

3.8. Rheumatoid arthritis genome-wide association study data set

Two seed genes (PTPN22 and IL2RA) generated suggestive evidence of association in the WTCCC-RA database (Supplemental Table 2). PTPN2 was previously associated with RA. We did not detect evidence

for association ($p < 0.05$) in any of the 26 network connecting genes in the WTCCC-RA data set after correcting for multiple testing within each gene.

3.9. Celiac Disease Genome-Wide Association Study data set

Four seed genes (IL12A, IL2, IL21, and SH2B3) showed significant association in the CeD data set (Supplemental Table 3). All four have been previously implicated in CeD. Two network connecting genes showed significant association to CeD including IL18RAP and NFATC2. Of these, NFATC2 has not been previously implicated in CeD. To further evaluate these findings, the NFATC2 SNP (rs228838) was genotyped in a new sample of CeD and showed significant association ($p < 0.0018$). This finding provides further evidence for the role of NFATC2, or a nearby gene, in CeD.

3.10. Systemic lupus erythematosus and multiple sclerosis genome-wide association study data set

Six of the nine seed genes were available for analysis in the SLE data set (PTPN22, IFIH1, STAT4, IL2, IL21, and IL2RA). All six were significantly associated with SLE in the present study. Of these six seed genes, four (IFIH1, IL2, IL21, and IL2RA) were not implicated in SLE at the time of our study but were implicated subsequent to our analysis. Two genes have been previously implicated (PTPN22 and STAT4). Of the six connecting genes showing evidence for association to SLE, five (IL12RB2, IL10RA, GRB2, RELB, and NFATC2) have not been previously implicated in SLE and therefore represent entirely novel genes (Supplemental Table 4). Only two of the nine seed genes were available for analysis in the MS data set (IL12A and IL2RA). Both were significantly associated with MS in the present study and other groups have also implicated both genes.

3.11. Analysis of candidate genes in a novel SLE case-control cohort

Network analysis identified six novel SLE candidate genes in the SLE study (Table 4). To further evaluate these findings, we genotyped all significant SNPs in the network connecting genes in a novel sample of 136 European American and 123 matched controls from the TRIDOM-SLE cohort study. Significant association to SLE was observed with IL12RB2 (rs56194793; $p = 0.045$) only. No association was observed with the rest of the SNPs in our small sample (data not shown).

3.12. Psoriasis genome-wide association study data set

Psoriasis (PSO) has been observed in patients who present with other AuD [21,39–41]. For instance, the prevalence of PSO in patients with CD is higher than expected by chance if they were mutually exclusive diseases [42,43]. Given the genetic and pathologic connections between psoriasis and other AuD, we hypothesized that our seed genes and candidate genes would be enriched for PSO. We therefore used a genome-wide psoriasis data set as a 'positive control' analysis. Table 4 lists the nine seed genes and the 26 connecting candidate genes. The PSO GWAS data set provides an independent analysis of the network candidate genes. IL21 was the only seed gene that showed significant association with PSO (rs12511287; $p < 0.002$) and has been recently implicated in PSO. We found evidence for association to PSO in four connecting gene (IL12RB2, REL, IL12B, and IL23A). IL12RB2 and REL were not implicated in PSO at the time of our study but were implicated subsequent to our analysis, therefore validating our approach.

Table 4
Association analysis in novel samples of 136 TRIDOM-SLE cases and 123 control subjects. The SNP showing strongest evidence for allelic association with SLE disease in the SLE GWAS data set was genotyped in a novel sample of 136 SLE cases and 123 control subjects.

SNP ID	Gene					*p*-value
rs56194793	IL12RB2	Alleles	CC	CT	TT	0.045
		Controls	6	27	83	
		Cases	10	40	74	

3.13. Summary of significant association results in seven autoimmune disorders

Fig. 3 presents a summary of the associated genes in at least two AuD and their overlap among the studied diseases. IL12RB2 and IL12B locus show associations in the WTCCC-CD, SLE, and PSO data sets (Supplemental Tables 1, 4, 5). A STAT3 locus is associated with CD and MS (Supplemental Tables 1 and 4). Similarly, the NFATC2 locus was associated with four autoimmune diseases independently. The association was detected in the WTCCC-CD, WTCCC-T1D, CeD, and SLE (Supplemental Tables 1, 2, 3, and 4).

3.14. Analysis of type-2-diabetes, bipolar, and autism disease, and alcoholism genome-wide association studies data sets

We evaluated the nine seed and 26 candidate connecting genes in non-AuD related samples from the WTCCC data sets, including the WTCCC-type 2 diabetes (WTCCC-T2D) and the WTCCC-Bipolar (WTCCC-BP) data sets (Supplemental Table 5). No association was observed for any of the nine seed genes in the WTCCC-T2D and WTCCC-BP data sets. Only one candidate gene (NFKB2) was significantly associated in the WTCCC-T2D data set (rs3802678 $p < 7.91 \times 10^{-6}$); and no association with any candidate genes in the WTCCC-BP data set (Supplemental Table 5). Furthermore, we evaluated the association of the candidate genes in GWAS of a case-control study of alcoholism, in which the subjects have been drawn from the Collaborative Study on the Genetics of Alcoholism [44] and Parkinson's disease [45] using the resources available at the HuGE and dbGaP (see Web-resources). Similar to WTCCC-BP and WTCCC-T2D, only one candidate gene in each GWAS showed association; the v-rel reticuloendotheliosis viral oncogene homolog A (RELA) and the suppressor of cytokine signaling 7 (COS7), respectively.

3.15. Linkage disequilibrium structure of candidate genes

We sought to examine the association among the identified SNPs within each gene in an effort to identify the portion of the associated gene with the phenotype, thus we examined the linkage disequilibrium (LD) among the significantly identified SNPs and their associated loci. NFATC2, IL12RB2, IL12B and STAT3 were the only genes/SNPs found to be significantly associated with at least two AuDs after Bonferroni correction. For NFATC2, the strongest association evidence was identified with WTCCC-CD (rs880324; $p < 0.004$), WTCCC-T1D (rs17199748; $p < 0.05$), CeD (rs228838: $p < 0.027$), and SLE (rs2869428: $p < 10^{-2}$-10^{-3}). All SNPs were located in different LD blocks and in very low LD (pairwise $r^2 > 0.5$) (data not shown). The association evidence for IL12RB2 was identified with WTCCC-CD (6 SNPs), NIDDK-CD (5 SNPs), SLE and PSO (rs11209032; $p < 3.2 \times 10^{-5}$, and rs11578380; $p < 0.017$). The rs11209032 revealed association in NIDDK-CD and Psoriasis data sets. Four SNPs of IL12B revealed significant association in the WTCCC-CD (rs1363670; $p < 0.009$, rs10042630; $p < 0.05$, rs6874870; $p < 0.037$, and rs2853696; $p < 0.039$), SNPs in the SLE, and 9 SNPs in Psoriasis data set (rs2082412; $p < 3.37 \times 10^{-7}$,10515780; $p < 0.026$, rs1549922; $p < 7.8 \times 10^{-5}$,rs2853694; $p < 3.7 \times 10^{-5}$, rs1433048; $p < 0.0007$, rs6894567; $p < 5.6 \times 10^{-6}$, rs4244437; $p < 0.00026$, rs983825; $p < 0.00083$, and rs1157509; 0.0001). The STAT3 gene on chromosome 17 reveled association to WTCCC-CD (rs744166; $p < 0.003$), SLE (rs3181224; $p < 10^{-4} - 10^{-5}$), and MS (rs3816769;

$p < 10^{-2} - 10^{-3}$) data set. Furthermore, 6 more SNPs in the STAT3 gene showed association to CD only (rs744166; $p < 0.0003$, rs1026961; $p < 0.0005$, rs7211777; $p < 0.0006$, rs4796793; $p < 0.03$, rs2883456; $p < 0.04$, rs12948909; $p < 0.008$, and rs1187180; $p < 0.013$). The common SNPs among the two data sets appear to be in one LD block (data not shown).

4. Discussion

In this study, we examined the hypothesis that network proximity within a whole genome molecular interaction network is a useful proxy for multigene patterns of inheritance. We examined nine genes predisposing to AuD using molecular interaction network analysis driven by IPA to build an AuD-specific network.

The first question we asked was whether the list of candidates "seed" genes map more proximal to one another than predicted by chance in molecular interaction databases. The molecular interaction networks used for this study capture on the order of fifty or more types of biological relationships, for examples: bind, phosphorylate, up-regulate, are co-expressed with, etc. Network proximity can presumably arise from shared biological properties or biological roles, but also from other, potentially confounding factors including: the order in which nodes (i.e, proteins) evolved [8,9], inherent 'scale-free' properties of biological networks (i.e., failure to correct for the contribution of hub nodes) [5,46], or other artifacts that can arise in the generation of the molecular networks. Clearly, the generation of false positive evidence for network proximity is a potential confound. In this study, we examined seven AuD using molecular interaction network analysis driven by IPA to build an AuD-specific network. The IPA analytics provide a statistic to estimate the likelihood that our observation could have occurred by chance ($p < 10^{-26}$). While this is an overwhelming significance value, the statistical algorithm was not developed specifically for our purposes and thus we could not be certain of its relevance. Reproducing the precise software parameters used for our test-case, we conducted 100 manual simulations with randomly generated sets of 9 seed genes. We have shown that all nine seed genes used in this study are directly connected to each other in a single 35-member molecular interaction network and that the interaction is significantly different from nine randomly chosen seed genes. Fig. 1 shows that the maximum number of seed genes connected to the default 35-member network ranges from 0 to 4 genes compared to our observed value of 9 AuD genes. The developed network is 35- interconnected genes; 9 seed genes surrounded by 26 network connecting genes. The cellular location of these network connecting genes and their association with various immune diseases (Fig. 2) support the hypothesis that they might harbor risk variants and are therefore strong "candidate genes" for AuD.

The fast emerging and publicly available GWAS datasets from a variety of sources provided an opportunity for us to test the association of our 26 candidate networking genes. Analysis of these genes shows that 16 generate suggestive to significant evidence of association to at least one AuD and in many cases more than one. This approach has identified genes already predicted to reside in regions established or recently reported to harbor risk variants in the seven AuD studied here.

Our rationale for making use of existing GWAS data is motivated by the fact that most studies typically report only highly significant associations (e.g. $p < 10^{-10}$) and consequently, truly associated variants with lower significance thresholds go unreported. Even the GWAS which lists the results of all published GWAS data only lists associations at 10^{-6} or lower. Thus, interrogating the raw GWAS data using innovative approaches such as network- and pathway-based analyses provides a powerful way to capture and report true disease associations. Several groups have adopted analytic and computational approaches to identify genetic risk factors for AuD. For example, De Jager et al., performed a meta-analysis across genotyping platforms using linkage disequilibrium-based allele imputation and identified several novel candidate genes for MS [47]. Goh et al., used integrated genetic data from the Online Mendelian Inheritance in Man (OMIM) database using a bipartite network-based visualization approach [48]. A similar approach was adopted by Baranzini using GWAS data from seven commonly known AuD. These studies identified groups of diseases that shared susceptibility genes and grouped them together, thus creating a disease landscape based on genetic similarity. Moreover, Baranzini showed that genes involved in one AuD also confer susceptibility to other AuD and specific candidate genes were identified and reported [49]. Our approach differs from these approaches because we used the seed-genes to create a network of proximal interacting genes "biological network proximity" within a whole genome molecular interaction network to infer candidate genes based on their physical topology.

A related, but more challenging question was whether the minimum number of genes (connecting genes) required to connect the maximum number of seed genes is enriched for genetic variation that predisposes individuals to AuD. The latter is challenging because there is no reliable way to distinguish false positive genetic association from real association when the genotype-phenotype correlation is relatively weak, as have been the vast majority of correlations for common genetic disorders to date. Our approach identified five genes that have not been previously associated with any of seven AuD and thus represent entirely novel genes for AuD etiology. These include IL10RA with SLE, DGKA with T1D, GRB2 with SLE, STAT5A with CD, and NFATC2 with CD, T1D, CeD and SLE. Several additional lines of evidence support a role for a subset of these novel genes in AuD. Of particular interest is the NFATC2 gene that codes for a protein belonging to the nuclear factor of activated T cells (NFAT) family that are involved in DNA binding with the REL-homology region (RHR) and an NFAT-homology region (NFR) [50]. NFATC2 is present in the cytosol and only translocates to the nucleus upon T cell receptor stimulation, where it becomes a member of the nuclear factors of activated T cells transcription complex and activates the well-established AuD gene IL2 [51]. This complex plays a central role in inducing gene transcription during the immune response. Notably, increased levels of NFATC2 have been shown to differentially regulate CD154 and IL-2 genes in T cells from patients with SLE [52] one of the disorders associated with NFATC2 in the present study. Importantly, we validated this finding using an independent sample set comprised of 136 patients with SLE and 123 matched controls from the TRIDOM-SLE cohort study.

Another novel gene of interest is IL10RA that was significantly associated with SLE in the present study. There have been no reported associations with IL10RA and SLE to date; however, IL10RA resides in a well-replicated risk region for CD on chromosome 11q23. Interestingly, a polymorphism in the mouse IL10RA gene has been associated with an SLE phenotype [53] thereby lending further support to our finding.

GRB2 was also of interest which we found associated with SLE. GRB2 codes for an adaptor protein associated with the MAPK pathway that is critical for maintenance of T cell tolerance, a process that is affected in patients with SLE. GRB2 in mice is known to interact with CD72, which is involved in B cell receptor signaling and a variety of early signaling events leading to activation of mitogen-activated protein (MAP) kinases. Deficiencies in CD72 leads to autoimmunity in mouse models and polymorphisms in this gene have been associated with several autoimmune disorders including SLE and T1D [54]. It has also been suggested that GRB2-related processes are altered in peripheral blood T lymphocytes from patients with SLE. Taken together, these studies support our finding that GBR2 is associated with SLE.

Our network analyses also identified three genes previously associated with at least one AuD that now we show associated with an entirely new AuD. These include IL12RB2 with SLE (previously associated with CD and psoriasis), NFKB1 with MS (previously associated with T1D and possibly CD), and EGFR with CD (previously associated with SLE). Furthermore, we provide numerous instances where our AuD-specific molecular networks identified candidate genes that were not known to be associated with AuD when our study was initiated but were later implicated by other groups. This provides strong support for

our network approach to identify novel AuD candidate genes. Examples of these associations include IL2, IL21, and IL12RB2 with CD, IL2RA with RA, IFIH1, IL2, IL2RA, and IL2RA with SLE, and IL21, IL12RB2 and REL with psoriasis. Finally, our analyses confirm associations with genes already known to be associated with AuD at the time our study, which further validates our approach. For instance, of the nine seed genes that were known to be associated with T1D at the time of our study, we found significant associations in six of these. Similarly, of the four seed genes associated with CeD at the time of our study, we found significant associations in all four.

Taken together, our study implicates a large number of previously reported as well as entirely novel candidate genes in a seven AuD. In many instance, we show that a single gene is associated with at least three or more AuD. Specifically, we demonstrate that four seed genes are implicated in at least three or more AuD: IL2 (CD, T1D, CeD, SLE), IL21 (CD, T1D, CeD, SLE, PSR), PTPN22 (T1D, RA, SLE) and IL2RA (T1D, RA, SLE). We also show that three candidate network genes are implicated in at least three or more AuD: IL12B (CD, SLE, PSR), NFATC2 (CD, T1D, CeD, SLE), and IL12RB2 (CD, SLE, and PSR).

Our 'negative control' analysis of the four non-AuD samples (T2D, Bipolar disorder and alcoholism) provide strong support for our rationale to use molecular networks to identify candidate genes because 1) none of the nine AuD seed genes were found to be associated with any of the four non-AuD samples and; 2) none of four network connecting genes that were associated with these non-AuD samples (NFKB2, LRRFIP1, RELA, and COS7) were found to be associated with any of the seven AuD sample. Interestingly, the RELA gene found to be associated with asthma in this study is critical for pulmonary host defense and has been associated with an asthma phenotype using cell lines and animal models [55–57].

A detailed LD structure for the candidate genes and the SNPs location were established to determine the extent of LD blocks and the exact location of each SNP. For example, for IL12RB2, the associated SNPs with AuD were found in two different LD blocks (Fig. 4). For the NFATC2, the associated SNPs were found in three different LD blocks (data not shown). This observation might give detailed insight about the various gene motifs and the role they play in various different but etiologically similar diseases. To confer risk association replication, we sought to genotype and test the association of the best-associated SNPs in a new AuD samples (TRIDOM). We only had sample sizes to test CD and SLE. Although modest, we showed replication of several SNPs within candidate genes to be associated with CD after Bonferroni adjustment. We also showed association to a small sample of SLE. To further evaluate the association of the candidate genes with the AuD, we are collecting large number of AuD samples for genotyping and analysis.

In this study, we have shown that molecular interaction networks for closely related, and previously identified nine AuD genes are highly connected more than could be explained by random chance. We also show that a network of connecting genes is also highly connected and the connecting genes could be treated as "candidate genes". The candidate genes examined in related phenotype databases showed significant enrichment for associated loci over non-immune-related databases. The network identifies a set of candidate genes with significant association, with some genes showing association to more than one phenotype. Even though it's beyond the scope of our expertise and interest, the development of a statistical metric to provide pre-analysis significant cut-off prediction would provide valuable for this type of analysis. The investigation of the particular molecular role of the identified genes in the molecular process of each of the AuD diseases or the overall AuD etiology warrants further investigation.

Disclosures

None.

Grant support

This work was supported by NIH grant numbers P30-HD-054.275 and RFA-HD-06.004 (TCG) and by generous donations from the following: David and Janice Katz, Judy McCormack, the Ryan Licht Sang Bipolar Foundation, and Debbie and Larry Hilibrand.

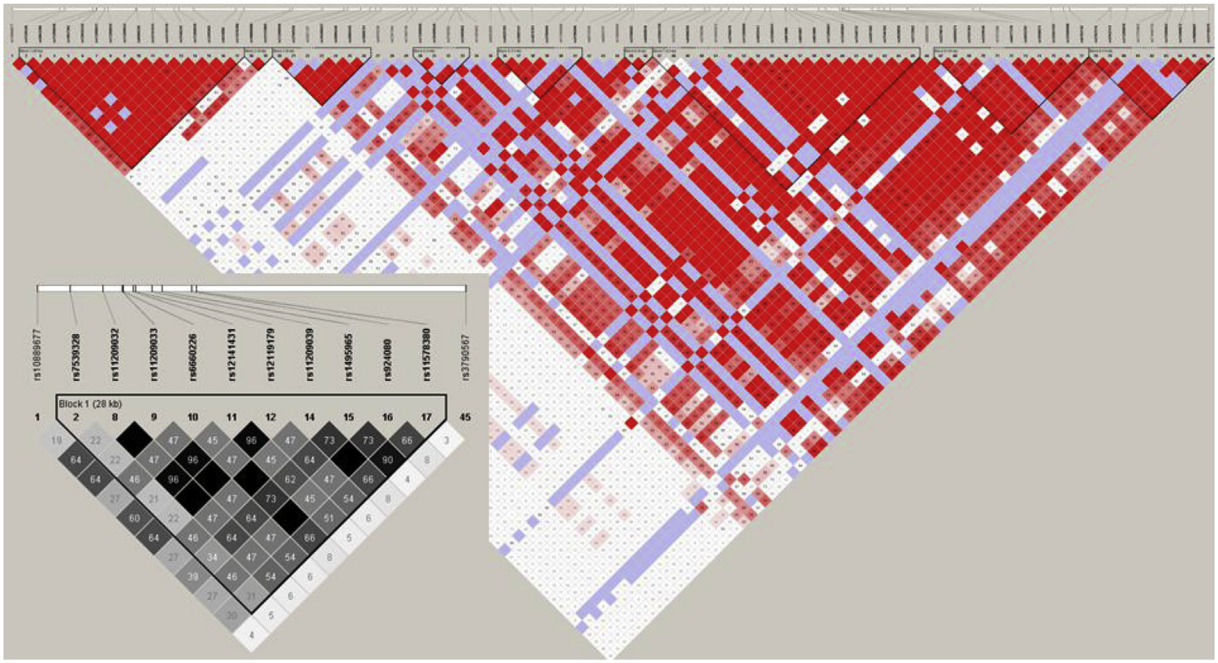

Fig. 4. The LD patterns of the IL12RB2 gene cluster region on chromosome 1 in population of European ancestry. The LD patterns (D′) were created in the Haploview and 1000 genome (See Materials and Methods) by using the genotyping data (only SNPs with MAF > 0.01). The indent represent the R^2 (significant SNPs) for the region harboring the significant SNPs genotyped in the WTCCC-CD (rs12119179, number 1), NIDDK-CD (rs10889677, number 3), and Psoriasis (rs11209032, number 8).

Author contributions

Sam Kara and Gerardo A. Pirela-Morillo analyzed the data and performed the bioinformatic analysis and the statistical support. Sam Kara, Gerardo A. Pirela-Morillo, and Conrad T. Gilliam drafted and revised the manuscript. Conrad T. Gilliam and George D. Wilson critically commented on the manuscript and provided important intellectual content. Sam Kara and Conrad T. Gilliam designed and supervised the project.

Acknowledgements

This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113.

The NIDDK IBDGC Crohn's Disease Genome-Wide Association Study was conducted by the NIDDK IBDGC Crohn's Disease Genome-Wide Association Study Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). This manuscript was not prepared in collaboration with Investigators of NIDDK IBDGC Crohn's Disease Genome-Wide Association Study and does not necessarily reflect the opinions or views of the NIDDK IBDGC Crohn's Disease Genome-Wide Association Study or the NIDDK.

The genotyping of Psoriasis samples was provided through the Genetic Association Information Network (GAIN). The data set(s) used for the analyses described in this manuscript were obtained from the database of Genotype and Phenotype (dbGaP; Web resources).

The genotyping of Multiple Sclerosis samples was provided by the National Institute of Neurological Disorders and Stroke (NINDS). The data sets used for the analyses described in this manuscript were obtained from the NINDS Database found at dbGap (Web resources).

Web resources

The URLs for data presented herein are as follows:

The Ingenuity Pathway Analysis (IPA) software (release 7.0): http://www.ingenuity.com.

The GeneOntology database (http://www.geneontology.org/):

The Pathway Studio 9 MammalPlus: http://www.pathwaystudio.com.

The Welcome Trust Case Control Consortium data set: http://www.wtccc.org.uk.

The dbGaP database which is maintained by the National Center for Biomedical Informatics: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap.

The Collaborative Study on the Genetics of Alcoholism: http://zork.wustl.edu/niaaa/

The GWAS associations and navigation software: http://hugenavigator.net/HuGENavigator/home.do.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jaut.2018.10.002.

References

[1] The Wellcome Trust Case Control Consortium & The Australo-Anglo-American Spondylitis Consortium (TASC), Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants, Nat. Genet. 39 (2007) 1329–1337.
[2] E.E. Schadt, S.A. Monks, T.A. Drake, A.J. Lusis, N. Che, V. Colinayo, et al., Genetics of gene expression surveyed in maize, mouse and man, Nature 422 (2003) 297–302.
[3] M.V. Rockman, L. Kruglyak, Genetics of global gene expression, Nat. Rev. Genet. 7 (2006) 862–872.
[4] V. Emilsson, G. Thorleifsson, B. Zhang, A.S. Leonardson, F. Zink, J. Zhu, et al., Genetics of gene expression and its effect on disease, Nature 452 (2008) 423–428.
[5] A.-L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization, Nat. Rev. Genet. 5 (2004) 101–113.
[6] C.C. Elbers, K.R. van Eijk, L. franke, F. Mulder, Y.T. van der Schouw, C. Wijmenga, et al., Using genome-wide pathway analysis to unravel the etiology of complex diseases, Genet. Epidemiol. 33 (2009) 419–431.
[7] M. Emily, T. Mailund, J. Hein, L. Schauser, M.H. Schierup, Using biological networks to search for interacting loci in genome-wide association studies, Eur. J. Hum. Genet. 17 (10) (2009) 1231–1240.
[8] A. Rzhetsky, I. Iossifov, T. Koike, M. Krauthammer, P. Kra, M. Morris, et al., GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data, J. Biomed. Inf. 37 (2004) 43–53.
[9] A. Rzhetsky, D. Wajngurt, N. Park, T. Zheng, Probing genetic overlap among complex human phenotypes, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 11694–11699.
[10] E.J. Rossin, K. Lage, S. Raychaudhuri, R.J. Xavier, D. Tatar, Y. Benita, et al., Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology, PLoS Genet. 7 (2011) e1001273.
[11] G. Lettre, J.D. Rioux, Autoimmune diseases: insights from genome-wide association studies, Hum. Mol. Genet. 17 (2008) R116–R121.
[12] D.M. Evans, J.C. Barrett, L.R. Cardon, To what extent do scans of non-synonymous SNPs complement denser genome-wide association studies? Eur. J. Hum. Genet. 16 (2008) 718–723.
[13] C.A. Anderson, G. Boucher, C.W. Lees, A. Franke, M. D'Amato, K.D. Taylor, et al., Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47, Nat. Genet. 43 (2011) 246–252.
[14] G.S. Cooper, F.W. Miller, D.R. Germolec, Occupational exposures and autoimmune diseases, Int. Immunopharm. 2 (2002) 303–313.
[15] S.A. Broadley, J. Deans, S.J. Sawcer, D. Clayton, D.A.S. Compston, Autoimmune disease in first-degree relatives of patients with multiple sclerosis: a UK survey, Brain 123 (2000) 1102–1111.
[16] G. Candore, D. Lio, G. Colonna Romano, C. Caruso, Pathogenesis of autoimmune diseases associated with 8.1 ancestral haplotype: effect of multiple gene interactions, Autoimmun. Rev. 1 (2002) 29–35.
[17] P.K. Gregersen, L.M. Olsson, Recent advances in the genetics of autoimmune disease, Annu. Rev. Immunol. 27 (2009) 363–391.
[18] T.R. Merriman, J.A. Todd, Genetics of autoimmune disease, Curr. Opin. Immunol. 7 (1995) 786–792.
[19] E.A.M. Festen, P. Goyette, T. Green, G. Boucher, C. Beauchamp, G. Trynka, et al., A Meta-Analysis of Genome-Wide Association Scans Identifies IL18RAP, PTPN2, TAGAP, and PUS10 As Shared Risk Loci for Crohn's Disease and Celiac Disease, PLoS Genet. 7 (2011) e1001283.
[20] V.E.H. Carlton, X. Hu, A.P. Chokkalingam, S.J. Schrodi, R. Brandon, H.C. Alexander, et al., PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid, Arthritis 77 (2005) 567–581.
[21] M. Uddin, M. Sturge, P. Rahman, M.O. Woods, Autosome-wide copy number variation association analysis for rheumatoid arthritis using the WTCCC high-density SNP genotype data, J. Rheumatol. 38 (5) (2011) 797–801.
[22] V. Orru, S.J. Tsai, B. Rueda, E. Fiorillo, S.M. Stanford, J. Dasgupta, et al., A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus, Hum. Mol. Genet. 18 (2009) 569–579.
[23] A. Petrone, M. Spoletini, S. Zampetti, M. Capizzi, S. Zavarella, J. Osborn, et al., The PTPN22 1858T gene variant in type 1 diabetes is associated with reduced residual {beta}-Cell function and worse metabolic control, Diabetes Care 31 (2008) 1214–1218.
[24] Vr Abadie, L.M. Sollid, L.B. Barreiro, B. Jabri, Integration of genetic and immunological insights into a model of celiac disease pathogenesis, Annu. Rev. Immunol. 29 (2011) 493–525.
[25] J. Gutierrez-Achury, R.C. de Almeida, C. Wijmenga, Shared genetics in celiac disease and other immune-mediated diseases, J. Intern. Med. 296 (6) (2011) 591–603.
[26] C. The International Multiple Sclerosis Genetics, Risk alleles for multiple sclerosis identified by a genomewide study, N. Engl. J. Med. 357 (9) (2007) 851–862.
[27] R.M. Plenge, M. Seielstad, L. Padyukov, A.T. Lee, E.F. Remmers, B. Ding, et al., TRAF1-C5 as a risk locus for rheumatoid arthritis – a genomewide study, N. Engl. J. Med. 357 (2007) 1199–1209.
[28] T. Gono, Y. Kawaguchi, T. Sugiura, T. Furuya, M. Kawamoto, M. Hanaoka, et al., Interferon-induced helicase (*IFIH1*) polymorphism with systemic lupus erythematosus and dermatomyositis/polymyositis, Mod. Rheumatol. 20 (2010) 466-470-70.
[29] D.J. Smyth, J.D. Cooper, R. Bailey, S. Field, O. Burren, L.J. Smink, et al., A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region, Nat. Genet. 38 (2006) 617–619.
[30] C. Winkler, C. Lauber, K. Adler, H. Grallert, T. Illig, A.-G. Ziegler, et al., An interferon-induced helicase (IFIH1) gene polymorphism associates with different rates of progression from autoimmunity to type 1 diabetes, Diabetes 60 (2011) 685–690.
[31] (IMSGC)* TIMSGC, IL12A, MPHOSPH9/CDK2AP1 and RGS1 are novel multiple sclerosis susceptibility loci, Genes Immun. 11 (2010) 397–405.
[32] K.A. Hunt, A. Zhernakova, G. Turner, G.A.R. Heap, L. Franke, M. Bruinenberg, et al., Newly identified genetic risk variants for celiac disease related to the immune response, Nat. Genet. 40 (2008) 395–402.
[33] A. Nikitin, E. Lavrikova, Y. Seregin, L. Zilberman, N. Tzitlidze, T. Kuraeva, et al., Association of the polymorphisms of the ERBB3 and SH2B3 genes with type 1 diabetes, Mol. Biol. 44 (2010) 228-32-32..
[34] A. Spurkland, L.M. Sollid, Mapping genes and pathways in autoimmune disease,

Trends Immunol. 27 (2006) 336–342.
[35] K. Wang, M. Li, M. Bucan, Pathway-based approaches for analysis of genomewide association studies, Am. J. Hum. Genet. 81 (2007) 1278–1283.
[36] Pailhez G, Bulbena A, Fullana MA, Castaño J. Anxiety disorders and joint hypermobility syndrome: the role of collagen tissue. Gen. Hosp. Psychiatr.;31:299-.
[37] S. Rafah, S.F. Beverly, N.K. Silvia, R. Lesley, A.M. Rachel, J. Meenakshi, et al., Genetic variation at the IRF7/PHRF1 locus is associated with autoantibody profile and serum interferon-alpha activity in lupus patients, Arthritis Rheum. 62 (2010) 553–561.
[38] D.R. Nyholt, A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other, Am. J. Hum. Genet. 74 (2004) 765–769.
[39] R.B. Warren, R.L. Smith, E. Flynn, J. Bowes, U. Consortium, S. Eyre, et al., A systematic investigation of confirmed autoimmune loci in early-onset psoriasis reveals an association with IL2/IL21, Br. J. Dermatol. 164 (2011) 660–664.
[40] R.P. Nair, K.C. Duffin, C. Helms, J. Ding, P.E. Stuart, D. Goldgar, et al., Genome-wide scan reveals association of psoriasis with IL-23 and NF-[kappa]B pathways, Nat. Genet. 41 (2009) 199–204.
[41] J. Bhalerao, A.M. Bowcock, The genetics of psoriasis: a complex disorder of the skin and immune system, Hum. Mol. Genet. 7 (1998) 1537–1545.
[42] D.J. Najarian, A.B. Gottlieb, Connections between psoriasis and Crohn's disease, J. Am. Acad. Dermatol. 48 (2003) 805–821.
[43] J.S. Danik, G. ParÃ©, D.I. Chasman, R.Y.L. Zee, D.J. Kwiatkowski, A. Parker, et al., Novel loci, including those related to Crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17 686 women/clinical perspective, Circ. Cardiovasc. Genet. 2 (2009) 134–141.
[44] H.J. Edenberg, D.M. Dick, X. Xuei, H. Tian, L. Almasy, L.O. Bauer, et al., Variations in GABRA2, encoding the [alpha]2 subunit of the GABAA receptor, are associated with alcohol dependence and with brain oscillations, Am. J. Hum. Genet. 74 (2004) 705–714.
[45] J. Simon-Sanchez, C. Schulte, J.M. Bras, M. Sharma, J.R. Gibbs, D. Berg, et al., Genome-wide association study reveals genetic risk underlying Parkinson's disease, Nat. Genet. 41 (2009) 1308–1312.
[46] A.-L. Barabasi, Scale-free networks: a decade and beyond, Science 325 (2009) 412–413.
[47] P.L. De Jager, X. Jia, J. Wang, P.I.W. de Bakker, L. Ottoboni, N.T. Aggarwal, et al., Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci, Nat. Genet. 41 (2009) 776–782.
[48] K.-I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, A-Ls BarabÃ¡si, The human disease network, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 8685–8690.
[49] S.E. Baranzini, The genetics of autoimmune diseases: a networked perspective, Curr. Opin. Immunol. 21 (2009) 596–605.
[50] M. Shimada, T. Miyagawa, M. Kawashima, S. Tanaka, Y. Honda, M. Honda, et al., An approach based on a genome-wide association study reveals candidate loci for narcolepsy, Hum. Genet. 128 (2010) 433-441-41.
[51] T.N. Nguyen, L.J. Kim, R.D. Walters, L.F. Drullinger, T.N. Lively, J.F. Kugel, et al., The C-terminal region of human NFATc2 binds cJun to synergistically activate interleukin-2 transcription, Mol. Immunol. 47 (2010) 2314–2322.
[52] V.C. Kyttaris, Y. Wang, Y.-T. Juang, A. Weinstein, G.C. Tsokos, Increased levels of NF-ATc2 differentially regulate CD154 and IL-2 genes in T cells from patients with systemic lupus erythematosus, J. Immunol. 178 (2007) 1960–1966.
[53] Z.-M. Qi, J. Wang, Z.-R. Sun, F.-M. Ma, Q.-R. Zhang, S. Hirose, et al., Polymorphism of the mouse gene for the interleukin 10 receptor alpha chain (Il10ra) and its association with the autoimmune phenotype, Immunogenetics 57 (2005) 697–702.
[54] H.-J. Wu, S. Bondada, CD72, a coreceptor with both positive and negative effects on B lymphocyte development and function, J. Clin. Immunol. 29 (2009) 12–21.
[55] L. Xiao, L. Jiang, Q. Hu, Y. Li, MiR-302e attenuates allergic inflammation in vitro model by targeting RelA, Biosci. Rep. 38 (2018).
[56] M. Rava, I. Ahmed, M. Kogevinas, N. Le Moual, E. Bouzigon, I. Curjuric, et al., Genes interacting with occupational exposures to low molecular weight Agents and irritants on adult-onset asthma in three european studies, Environ. Health Perspect. 125 (2017) 207–214.
[57] Y. Zhou, G.F. Wang, L. Yang, F. Liu, J.Q. Kang, R.L. Wang, et al., Treatment with 1,25(OH)2D3 induced HDAC2 expression and reduced NF-kappaB p65 expression in a rat model of OVA-induced asthma, Braz. J. Med. Biol. Res. Revista brasileira de pesquisas medicas e biologicas 48 (2015) 654–664.