

Research Article

The Autoimmune Tautology: An In Silico Approach

Ricardo A. Cifuentes,¹ Daniel Restrepo-Montoya,² and Juan-Manuel Anaya¹

¹ Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24, No. 63-69 piso 3, Bogotá, Colombia

² Bioinformatics and Intelligent Systems Research Laboratory (BIOLISI), Universidad Nacional, Avenida Carrera 30, No. 45-03, Bogotá, Colombia

Correspondence should be addressed to Ricardo A. Cifuentes, ricardo.cifuentes@urosario.edu.co

Received 13 October 2011; Accepted 26 November 2011

Academic Editor: Adriana Rojas-Villarraga

Copyright © 2012 Ricardo A. Cifuentes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is genetic evidence of similarities and differences among autoimmune diseases (AIDs) that warrants looking at a general panorama of what has been published. Thus, our aim was to determine the main shared genes and to what extent they contribute to building clusters of AIDs. We combined a text-mining approach to build clusters of genetic concept profiles (GCPs) from the literature in MedLine with knowledge of protein-protein interactions to confirm if genes in GCP encode proteins that truly interact. We found three clusters in which the genes with the highest contribution encoded proteins that showed strong and specific interactions. After projecting the AIDs on a plane, two clusters could be discerned: Sjögren's syndrome—systemic lupus erythematosus, and autoimmune thyroid disease—type 1 diabetes—rheumatoid arthritis. Our results support the common origin of AIDs and the role of genes involved in apoptosis such as *CTLA4*, *FASLG*, and *IL10*.

1. Introduction

There are clinical and genetic grounds for assuming similar immunogenetic mechanisms in autoimmune diseases (AIDs). Clinical evidence highlights the cooccurrence of distinct AIDs within members of a nuclear family and within an individual [1]. Individuals with a multiple autoimmune syndrome (MAS) have been grouped into three basic groups in which various AIDs cluster around one of three “main” AIDs, namely, systemic lupus erythematosus (SLE), autoimmune thyroid disease (AITD), and primary Sjögren's syndrome (SS). These three might be considered the “chaperones” of the other AID [2]. Along the same line of clinical evidence, there are therapies such as tumor necrosis factor inhibitors, rituximab, or a gluten-free diet that are already proving effective for more than one AID [3, 4]. With regards to genetic evidence, it has also been stated that around 44% of the single nucleotide polymorphisms (SNPs), which were found in genome-wide association studies (GWAS) on AIDs, are shared by two or more of the following diseases: celiac disease, Crohn's disease, psoriasis, multiple sclerosis (MS),

rheumatoid arthritis (RA), type 1 diabetes (T1D), and SLE [5].

There are also genetic differences among AIDs. In spite of sharing several susceptibility genes, the differences among most AIDs, in particular systemic ones such as SLE and RA, seem to reside in the contribution of each gene to each disease [6]. Additionally, clusters of AIDs have been described where SNPs that make an individual susceptible to one class of AIDs also protect from another class of AIDs [7]. Furthermore, it is already known that different AIDs are associated with some different alleles from the human leukocyte antigen (HLA) [6].

As a consequence, it is important to obtain a general panorama of the problem in order to understand the origin of the AIDs. However, in biomedical research, the amount of experimental data and published scientific information is overwhelming. Therefore, literature-based discovery (LBD) tools emerge as useful to make the biomedical literature accessible for research purposes [8]. Thus, different LBD methods have been used to mine large amounts of literature and find the necessary information (Table 1) [8–11] with

TABLE 1: Examples of literature-based knowledge discovery tools.

Tool	Mined data	URL
ANNI	MedLine	http://www.biosemantics.org
Arrowsmith ¹ ,	MedLine, OVID	http://wiki.uchicago.edu/
	UMLS concepts in	
Arrowsmith ²	title words (MedLine)	http://arrowsmith.psych.uic.edu/
BITOLA	MeSH and LocusLink	http://www.mf.uni-lj.si/bitola/
LitLinker	UMLS	http://litlinker.ischool.washington.edu/
FACTA	MedLine	http://refine1-nactem.mc.man.ac.uk/facta/
FAUN	MedLine	https://grits.eecs.utk.edu/faun/

¹ University of Chicago

² University of Illinois at Chicago

For more information about biomedical text mining tools visit
http://arrowsmith.psych.uic.edu/arrowsmith_uic/tools.html.

two main approaches in the biomedical domain [12]. One approach focuses on the extraction of precise relationships between concepts, and the other relates biomedical concepts one to each other based on the statistical properties of their occurrence and cooccurrence in literature. A known LBD method based on concept occurrence is the concept profile (CP), in which a concept is characterized by a list of associated concepts, together with weights that indicate the strength of the association [13].

The output of the concept profiling method is a list of associations ordered by the strength of their relationship that needs verification. It is typically done with domain-relevant knowledge usually based on expert human judgments or even experimental validation [8, 14]. The latter approach is currently more feasible in the biomedical field given the increase in experimentally identified binary interactions between proteins that has made it possible to see how these components come together to form large functional regulatory networks [15]. There are several network approaches [16] that could be organized based on the type of biological or molecular interactions [17] and that analyze diverse databases (Table 2) [18–24]. Thus, the information related to protein-protein interactions helps us to study these associations from the perspective of biochemistry, signal transduction, and biomolecular networks [25]. Identification of functional roles of unknown pathogenic genes can also make it possible to understand pathogenic mechanisms. Proteins that are tightly connected in biological networks often work in similar processes [26].

This complex panorama shows that we are still distant from knowing everything, that is to know about genes, their interactions with other genes, and their impact on biological functions [6]. Therefore, the aim of this study was to obtain information from the literature and annotated databases to find main common genes in autoimmunity and determine to what extent they contribute to different clusters of AIDs.

2. Methods

Our analysis was made by using experimental knowledge of protein-protein interaction to evaluate the top ranked genes,

which had been found through the CP approach to mine the biomedical literature (Figure 1).

2.1. Literature-Based Knowledge Discovery. The concepts selected as input for the LBD software were the three AIDs referred to as chaperones of autoimmunity (i.e., AITD, SS, and SLE). We also selected as input concepts the AIDs mentioned in literature as present in relatives of probands of these three diseases: MS, RA, T1D, vitiligo (VIT), and systemic sclerosis (SSc) [2].

To evaluate the genetic similarity of those AIDs, we chose the Anni software because it uses the CP methodology that has proven to be effective for finding information in the form of associations in the biological domain [27]. First, the mapping of those concepts in the thesaurus of the Anni software that uses the concept profile methodology was evaluated [28]. At this point, we eliminated the VIT concept because it showed ambiguity in mapping. Next, the CP for each one of the seven remaining AIDs was built. Those profiles corresponded to the weighted list made by all the genes mentioned in MedLine, so they were called genetic CPs (GCPs). To do this, we selected the 25,010 genes that belong to human beings from the thesaurus in Anni, and, then, we mined all the MedLine records that contained these genes in their text. Next, the associations between GCP were explored through hierarchical clustering. The clusters were generated by matching the GCP for each one of the mapped AIDs, as the CP can be described as vectors. Then, the similarities between the GCP in the found clusters were analyzed. For this purpose, we obtained a cohesion score by using as an inclusive filter for matching the described 25,010 genes. Briefly, the cohesion score is an average of the inner products of all possible pairs of profiles corresponding to the concepts in the group of interest. The contribution of each gene in the profile to the cohesion score was assessed in terms of percentage. To interpret the cohesion score we used a *P* value that gives the probability that the same score or higher would be found in a random group of the same size. This *P*-value was obtained by using the default parameter in Anni of 200 iterations. Finally, the distances between concepts that reflect the matching value between GCPs were projected in

TABLE 2: Examples of tools to analyze biological pathways.

Tool	Analyzed data	URL
Cytoscape	220 diverse databases.	http://www.cytoscape.org/
BIANA	uniprot, GenBank, IntAct, KEGG and PFAM.	http://sbi.imim.es/web/BIANA.php
Pathway studio	MedLine.	http://www.ariadnegenomics.com/products/pathway-studio/expression-analysis/algorithms
Patika	Reactome, UniProt, Entrez Gene, and GO.	http://www.patika.org/
Genes2networks	BIND, DIP, IntAct, MINT, pdzbase, SAVI, Stelzl, vidal, ncbi hprd, and KEGG mammalian	http://actin.pharm.mssm.edu/genes2networks/

Verification of literature-based discovery with protein-protein interaction knowledge in autoimmune diseases

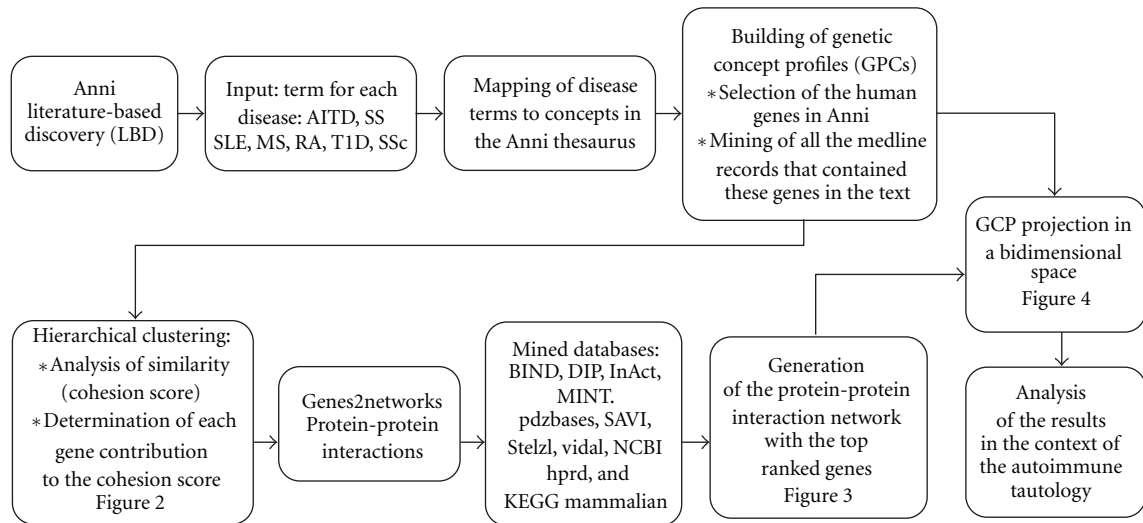


FIGURE 1: Flowchart of the methodology. AITD: autoimmune thyroid disease, SS: primary Sjögren's syndrome, SLE: namely systemic lupus erythematosus, MS: multiple sclerosis, RA: rheumatoid arthritis, T1D: type 1 diabetes, and SS: systemic sclerosis.

a two-dimensional space, in order to understand the AID clustering.

2.2. Network Analysis. To analyze if the genes in the clusters previously found through LBD corresponded to proteins with a known interaction, a network analysis was done with the genes that contributed at least 0.1% to any of the clusters found by the method described in Section 2.1. For this purpose, the software, Genes2networks, was selected because it finds relationships between proteins by using ten high quality mammalian protein-protein interaction databases that take into account not only filtered high throughput but also low throughput experiments that have a lower probability of false positives [29]. Then, in order to find tightly connected proteins, the settings that were used in Genes2networks to build the network were (1) no filter for minimum number of references, (2) the maximum links per reference were four, (3) a maximum pathway length of two,

and (4) a significant Zscore of 2.5 of the intermediate nodes, which was calculated through a binomial proportions test, as previously described [29].

2.3. Systematic Search. We did a classical systematic search, as previously done by our group [30], to understand the relevance of the genes found by our approach on AIDs. The genes selected were ones that contributed more than 1% to two or more clusters of AIDs and were close to each other in subnetworks where they were separated by a maximum of one node. To do this, we did a systematic search of the Catalog of Published Genome-Wide Association Studies at <http://www.genome.gov/26525384> and on PubMed by using three terms: the gene name, the MeSH term “genome-wide association study” and the MeSH term for each AIDs that belonged to the found clusters. Consequently, the terms for the AIDs were chosen from the next MeSH terms: “arthritis, rheumatoid,” “multiple sclerosis,” “diabetes mellitus, type 1,”

TABLE 3: Genes with a contribution higher than 0.1% to the found clusters of the studied autoimmune diseases.

Cluster 1. SLE-SS		Cluster 2. T1D-AITD		Cluster 4. RA-MS	
Gene	%	Gene	%	Gene	%
<i>TRIM21</i>	27.91	<i>TPO</i>	32.4	<i>TNF</i>	39.5
<i>TNFSF13B</i>	27.46	<i>CTLA4</i>	28.6	<i>HLA-DRB1</i>	20.7
<i>TROVE2</i>	19.8	<i>TNFRSF25</i>	6.7	<i>IL10</i>	5.2
<i>SSB</i>	6.6	<i>HLA-DRB1</i>	6.7	<i>IL6</i>	2.2
<i>FAS</i>	2.7	<i>PTPN22</i>	6.4	<i>CCL2</i>	0.6
<i>DLAT</i>	2.6	<i>GAD1</i>	4.6	<i>CD4</i>	0.6
<i>IRF5</i>	1.0	<i>GAD2</i>	3.6	<i>MMP9</i>	0.6
<i>IL10</i>	0.9	<i>AIRE</i>	1.7	<i>IL1B</i>	0.5
<i>FASLG</i>	0.8	<i>PTPRN</i>	1.5	<i>IL4</i>	0.5
<i>TNFRSF25</i>	0.6	<i>HLA-DQB1</i>	0.5	<i>TNFSF13B</i>	0.5
<i>CR1</i>	0.5	<i>IDDM2</i>	0.5	<i>IL23A</i>	0.4
<i>CALR</i>	0.5	<i>SUMO4</i>	0.5	<i>CCR2</i>	0.4
<i>SPTAN1</i>	0.4	<i>ICA1</i>	0.4	<i>IL1RN</i>	0.4
<i>RNPC3</i>	0.4	<i>FOXP3</i>	0.3	<i>CCL5</i>	0.3
<i>CR2</i>	0.2	<i>FCRL3</i>	0.2	<i>ICAM1</i>	0.3
<i>SNRNP70</i>	0.2	<i>CD4</i>	0.2	<i>CXCR3</i>	0.3
<i>SERPIND1</i>	0.2	<i>FASLG</i>	0.2	<i>HLA-DQB1</i>	0.3
<i>C1QA</i>	0.2	<i>CXCL10</i>	0.2	<i>VCAM1</i>	0.2
<i>IL18</i>	0.2	<i>CD8A</i>	0.2	<i>CTLA4</i>	0.2
<i>IL6</i>	0.2	<i>IL1B</i>	0.2	<i>PADI4</i>	0.2
		<i>TSHR</i>	0.2	<i>IFNB1</i>	0.2
				<i>CRP</i>	0.2
				<i>CCR5</i>	0.2
				<i>IL12B</i>	0.2

SLE: systemic lupus erythematosus, SS: Sjögren's syndrome, T1D: type 1 diabetes, AITD: autoimmune thyroid disease, RA: rheumatoid arthritis, MS: multiple sclerosis, %: percentage of contribution to the cluster.

“lupus erythematosus, systemic,” “scleroderma, systemic” and “Sjögren's syndrome.” In the case of thyroid disease, the term “thyroid” was used. The information from PubMed was excluded when the retrieved information did not explicitly refer to the specific gene, for instance when CD4 referred to a type of cell (i.e., lymphocyte) but not to the gene.

3. Results

There were three paired clusters with a probability equal to or less than 3 percent that their cohesion score would be found in a random group of the same size: SLE with SS ($P = 0.02$), T1D with AITD ($P = 0.02$), and RA with MS ($P = 0.03$) (Figure 2). Regarding the genes that contributed to building the clusters, 55 of them had a contribution higher than 0.1% to the cohesion score of any of those clusters. Some of them were shared by more than one cluster: *HLA-DQB1*, *CD4*, *TNFSF25*, *FASLG*, *IL1B*, *IL6*, *IL10*, *TNFSF13B*, *CTLA4* and *HLA-DRB1*. The later three had a contribution higher than 20% to any of the three specific clusters. The other genes contributed to only one cluster. It should be mentioned that there were also specific genes for one cluster that had a contribution of around 20% to their clusters, such

as *TRIM21* and *TROVE2* in the cluster made up of SLE and SS, *TPO* in the cluster made up of T1D and AITD, and *TNF* in the cluster made up of RA and MS (Table 3).

Concerning to the network analysis, we used as input the previously mentioned 55 genes. 29 of these 55 entries were identified and described on the graph (Figure 3). Some genes such as *IL6* and *HLA-DRB1* did not appear in the network. This could have been because of the strict threshold, a maximum pathway length of two, established to avoid weak interactions or because they did not have protein-protein interactions already reported in the used database. For instance, some genes relating to antigen presentation such as *HLA-DRB1* may be absent in protein interaction networks.

The network had 20 intermediary nodes, 19 significant with a Z score above the cutoff of 2.5 (Table 4), thus indicating that they may be specific to interact with components from the inputted seed list of genes. In other words, those results indicated that the seed genes encode proteins that had strong and specific interactions. In the graph, it can also be seen that the genes common to more than one cluster belonged to the same connected network (Figure 3). There were two subnetworks of genes that had

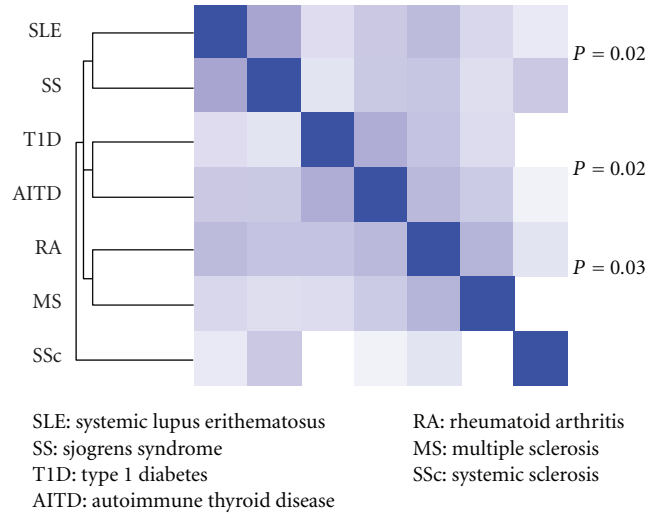


FIGURE 2: Clustering of seven autoimmune diseases. SLE: systemic lupus erythematosus, SS: Sjögren's syndrome, T1D: type 1 diabetes, AITD: autoimmune thyroid disease, RA: rheumatoid arthritis, MS: multiple sclerosis, SSc: systemic sclerosis.

TABLE 4: Significance of intermediates sorted by z-score.

Gene name	Link	Link in background	Links to seed	Links in subnetwork	z-score
HLA-DQA2	3	11429	2	60	15,852
DARC	4	11429	2	60	13,692
LCK	67	11429	6	60	9,548
PRTN3	9	11429	2	60	9,007
APCS	10	11429	2	60	8,522
FN1	62	11429	5	60	8,215
IGFBP7	11	11429	2	60	8,103
PTPN13	12	11429	2	60	7,737
CASP1	18	11429	2	60	6,215
A2M	24	11429	2	60	5,293
DCN	25	11429	2	60	5,171
NCL	30	11429	2	60	4,655
C3	31	11429	2	60	4,566
JAK2	116	11429	4	60	4,356
PTPRC	35	11429	2	60	4,248
THBS1	37	11429	2	60	4,108
ARRB1	44	11429	2	60	3,690
TRADD	63	11429	2	60	2,910
PIK3R1	133	11429	3	60	2,761
FYN	153	11429	3	60	2,457

a contribution higher than 0.1% and that were shared by more than one cluster. The first was made up of *HLA-DQB1*, *CD4*, *CTLA4* and *FASLG* that were genes connected through only one internode (*TNFRSF25* is also connected through three internodes with *FASLG*) and the second subnetwork was made up of *IL1B* and *IL10* that was connected to *TNF*, the gene with the highest contribution to the cluster made by RA

and MS. There was also another subnetwork made with the directly connected *CIQA*, *CR1*, and *CR2* genes that belonged to the cluster made by SLE and SS (Figure 3).

We also observed that some of the genes with a contribution higher than 0.1% to only one cluster belonged to three little separate networks. The first little network had the genes *GAD1* and *GAD2* from the cluster of T1D-AITD, the second

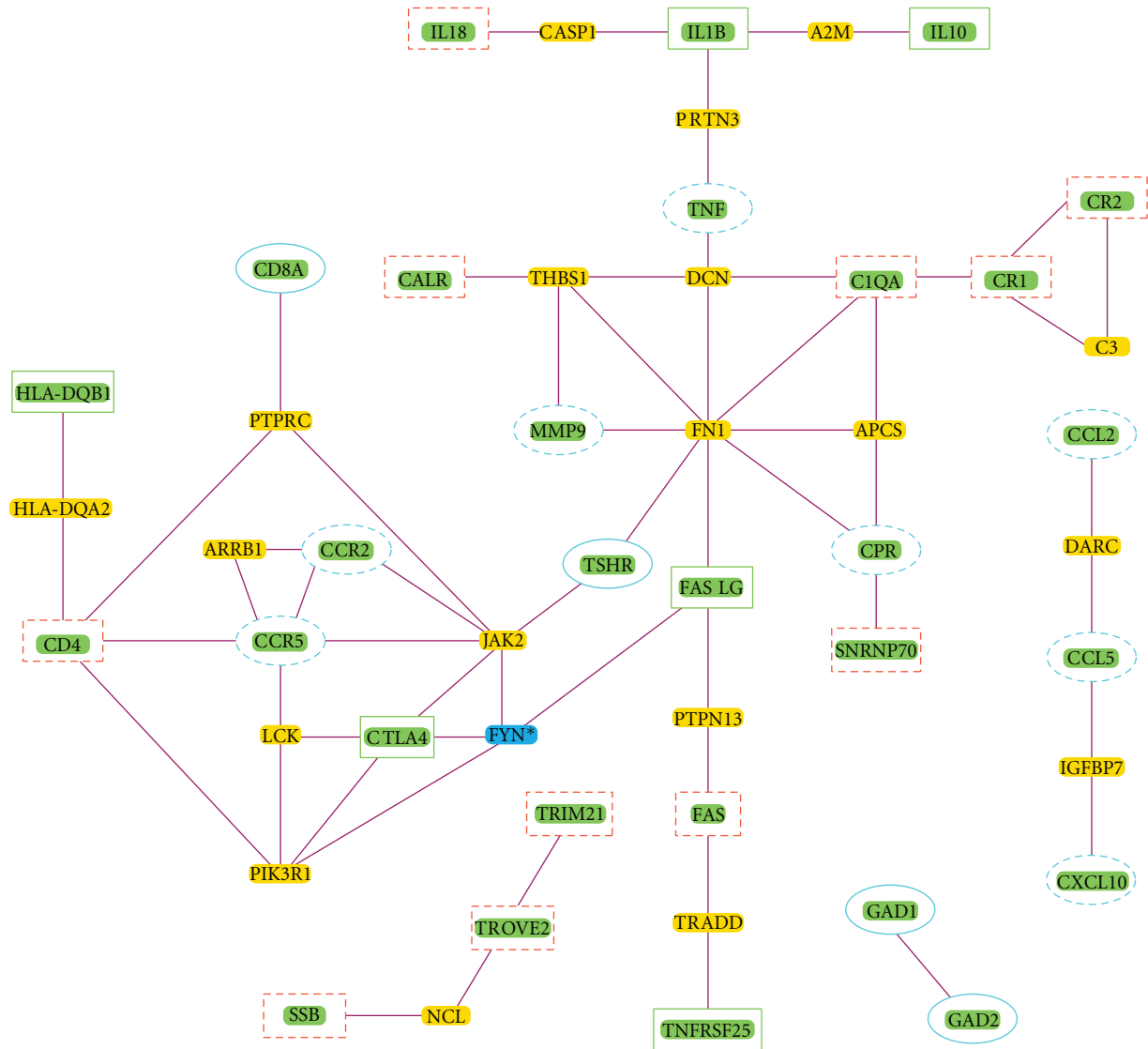


FIGURE 3: Network analysis of the genes that contribute to the clusters of autoimmune diseases. Solid squares: genes with a contribution higher than 0.1% that are shared by more than one cluster. Dotted squares: genes with a contribution higher than 0.1% from the SLE-SS cluster. Solid ovals: genes with a contribution higher than 0.1% from the T1D-AITD cluster. Dotted ovals: genes with a contribution higher than 0.1% from the RA-MS cluster. The other nodes correspond to significant intermediary ones (the asterisk indicates a nonsignificant intermediary node).

had the genes *TRIM21*, *TROVE2*, and *SSB* from the cluster of SLE-SS, and the third had the genes *CCL5* and *CCL2* from the cluster RA-MS (Figure 3).

Through the systematic search, we looked for GWAS information on six genes (Table 5). *HLA-DQB1* [31], *CTLA4* [32, 33], and *FASLG* and *IL10* [34] were related to AIDs in GWAS. In contrast, to date *CD4* and *IL1B* have not been related by GWAS data to any of the above-mentioned AIDs.

Finally, according to the distances obtained through the LBD approach, the evaluated AIDs were projected into two main spaces that are near each other. The first included SS and SLE, and the second, AITD, T1D, and RA. Both were distant from SSc and a little closer to MS, especially in the case of the RA (Figure 4).

4. Discussion

Our *in silico* approach that combined LBD and network analysis of protein-protein interactions allowed us to confirm common genes involved in autoimmunity as well as to estimate their contribution into the clusters of AIDs. Some common genes made an important contribution to only one specific cluster such as *TRIM21*, *TROVE2*, or *SSB*, but others were present in more clusters of AIDs such as *HLA-DQB1*, *FASLG*, *CTLA4*, or *CD4*. However, our approach did not intend to find all the genes shared among AIDs. In fact, not all the genes could be validated through protein-protein interactions, and others did not make a significant contribution to the described clusters of AIDs.

TABLE 5: Relevance on autoimmunity GWAS of the genes with a contribution higher than 1% to two or more clusters of the studied autoimmune diseases.

Gene	Full name	Location	GWAS catalogue	Reference
<i>HLA-DQB1</i>	Major histocompatibility complex, class II, DQ beta 1	6p21.3	MS, PBC, RA, SSc, CD, UC, CrD	[31]
<i>CD4</i>	CD4 molecule	12pter-p12	—	—
<i>CTLA4</i>	Cytotoxic T-lymphocyte-associated protein 4	2q33	T1D, RA, MS, SLE, CD	[32, 33]
<i>FASLG</i>	Fas ligand (TNF superfamily, member 6)	1q23	CD, CrD	—
<i>IL1B</i>	Interleukin 1, beta	2q14	—	—
<i>IL10</i>	Interleukin 10	1q31-q32	T1D, SLE, UC, CrD	[34]

MS: multiple sclerosis, PBC: primary biliar cirrhosis, RA: rheumatoid arthritis, SSc: systemic sclerosis, CD: celiac disease, CrD: crohn disease, T1D: Type 1 diabetes, SLE: systemic lupus erithematosus, UC: ulcerative colitis, PSO: Psoriasis.

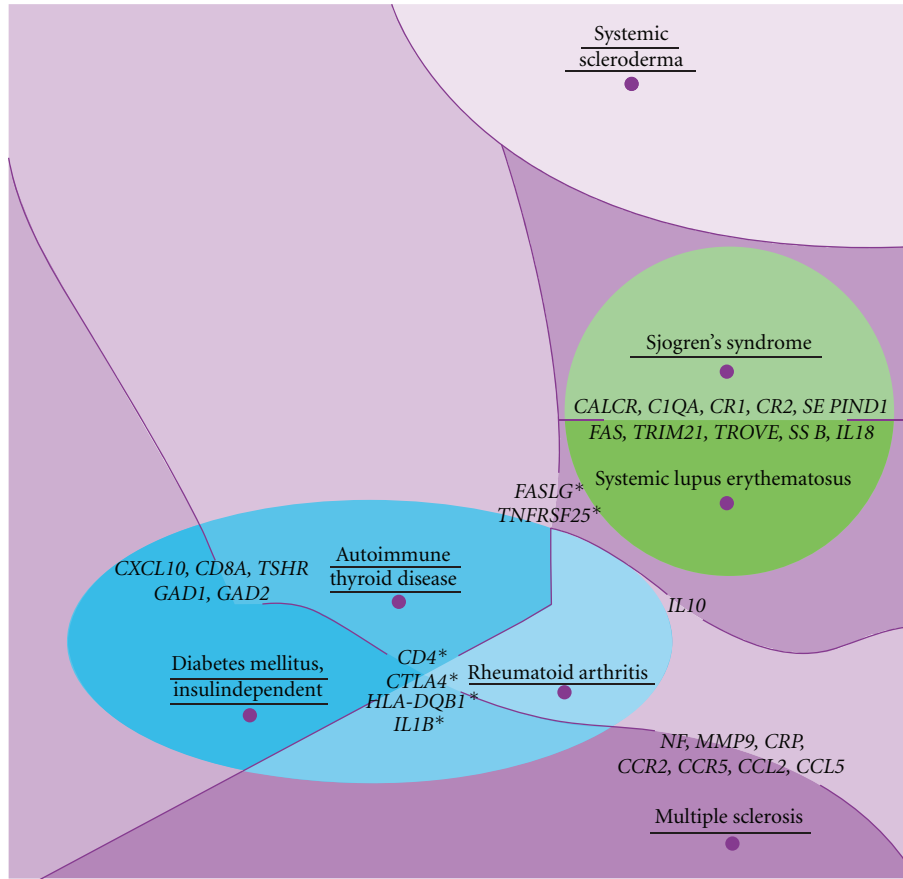


FIGURE 4: Projection of the seven studied autoimmune diseases on a plane. This figure shows the shared space of the genetic concept profiles from the studied AIDs (underlined), according to the matching value of their genetic concept profiles. We can see the genes with a contribution to clustering higher than 0.2%, the asterisk indicates the genes shared by two clusters.

With regards to genes shared by more than one cluster of AIDs, it can be seen that they were typically found to be significant in GWAS. However, there were exceptions. In the case of *CD4*, an association was not found with any AID by GWAS, but another approach that combines biological similarities found that *CD4* is a likely causal gene of RA [35], one that had been seen as high risk by recent studies [36, 37]. In contrast to GWAS, the genes that were found to be related to RA by the approach that combines biological similarities

could be easily classified into related functional categories or biological processes [35], thus making these finding similar to our results.

In contrast, there were genes that contributed mainly to specific clusters of AIDs such as *TRIM21* (*Ro52*), *TROVE2* (*Ro60*) and *SSB* (*La*) that were found to be important for the SLE-SS cluster. In spite of the fact that they were not significant at the GWAS level, this observation agreed with the fact that anti-SS-A (*Ro52/Ro60*) autoantibodies have

been described as serological markers for both SS and SLE [38–40]. Ro52 works as an E3 ligase and mediates ubiquitination of several members of the interferon regulatory factor (IRF) family. Its deficiency has been associated with enhanced production of proinflammatory cytokines that are regulated by the IRF transcription factors including cytokines involved in the Th17 pathway [41]. Although Ro ribonucleoproteins such as Ro60 and La were discovered many years ago, their function is still poorly understood [42]. It has been suggested that *TROVE2* acts as a modulator in the Y RNA-derived miRNA biogenesis pathway. The hypothesis is that Ro RNPs are “latent” pre-miRNAs that can be converted into miRNAs under certain circumstances [42]. In addition, it was observed that narrow-band ultraviolet B irradiation provoked significant alterations of the keratinocyte morphology and led to the membrane expression of antigens recognized by anti-La and anti-Ro 60 kDa sera [43].

Another observation about genes that contributed mainly to specific clusters was that genes typically involved in one AID such as *C1QA* and *CR1* in the case of SLE, or *GAD1* and *GAD2* in the case of T1D, were found by our approach to be shared with SS or AITD, respectively. These findings agree with the observations that around 24% of patients with T1D expressed antithyroid autoantibodies and that 17% of them had AITD in comparison to 6% of age-matched controls [44].

The projection of the AIDs on a plane agreed with the similarity between genetic variation profiles of T1D and AITD found by another approach, which builds genetic variation profiles taking into account *P* values and odds-ratios of significant SNPs in GWAS, but does not totally agree with the claimed opposition between MS and RA [7]. It can be seen that RA has some similarity with MS in spite of being closer to AITD. This projection also agreed with the behavior of HLA, even in admixed Latin-American populations, as diseases that were closer in it shared risk alleles. This is the case for SLE, SS, and T1D that have the *DRB1*03:01* allele as a risk factor [30, 45, 46]. Furthermore, in diseases that are distant in our clustering analysis, such as MS and T1D, the same *DQB1*06:02* allele gives protection to the first but risk to the second disease [47].

From the biological perspective, our results showed the central role of *FASLG* as it is connected through one node to *CTLA4*, which is connected to *CD4* through one node and that, in turn, is connected to *HLA-DQB1* the same way (Figure 3). *FASLG* is also connected with *TNF* through two nodes, and this is connected, in turn, through one node to *IL1B*, which is also connected through one node to *IL10* and *IL18*. It is interesting that these two pathways are involved in similar processes since *CTLA4*, and *IL10* are implicated in peripheral immunologic tolerance [48]. *FASLG* is also connected to two other pathways. It is connected through one node to *C1QA*, which is directly connected to *CR1*. Lastly, it is also indirectly connected to the pathway of *TROVE2*, *TRIM21*, and *SSB* through a route that was not shown on the graph. This route involved *SUMO1*, a gene that has been associated with a blockage of the FAS pathway

in RA, thus preventing apoptosis [49]. Taken together, our results highlight the autoimmunity role of genes involved in the process of apoptosis such as *CTLA4*, *FASLG*, and *IL10* that work together with genes involved in the inflammatory process such as *IL1B* [50].

Biomedical informatics involves a core set of methodologies that can provide a foundation for crossing the “translational barriers” associated with translational medicine [51]. Since the classical systematic review of literature could be incomplete because a significant amount of conceptual information present in literature is missing from the manually indexed terms [10], it seems to be advisable to combine the classical approach for searching literature with these new techniques.

In summary, the bioinformatics approach that combines text mining and network analysis of proteins allowed functional modules of interacting disease genes to be identified and can be used to predict additional disease gene candidates. Our approach also gave further evidence of the common origin of AIDs as the clustering of these diseases took into account thousands of genes that contribute to make the genetic concept profiles. Furthermore, this mining approach identified the specific contribution of a number of genes to causing some AIDs to cluster. These genes could be useful for further research.

Abbreviations

AIDs:	Autoimmune diseases
AITD:	Autoimmune thyroid disease
CP:	Concept profile
GCP:	Genetic concept profiles
GWAS:	Genome-wide association studies
HLA:	Human leukocyte antigen
IRF:	Interferon regulatory factor
LBD:	Literature-based discovery
MAS:	Multiple autoimmune syndrome
MS:	Multiple sclerosis
RA:	Rheumatoid arthritis
SLE:	Systemic lupus erythematosus
SNPs:	Single nucleotide polymorphisms
SS:	Primary Sjögren’s syndrome
SSc:	Systemic sclerosis
T1D:	Type 1 diabetes
VIT:	Vitiligo.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgments

The authors are grateful to the members of the Center for Autoimmune Diseases Research (CREA) for fruitful discussions. This work was supported by the School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia.

References

- [1] J. M. Anaya, "The autoimmune tautology," *Arthritis Research & Therapy*, vol. 12, no. 6, p. 147, 2010.
- [2] J. M. Anaya, R. Corena, J. Castiblanco, A. Rojas-Villarraga, and Y. Shoenfeld, "The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity," *Expert Review of Clinical Immunology*, vol. 3, no. 4, pp. 623–635, 2007.
- [3] J. Gutierrez-Achury, R. Coutinho de Almeida, and C. Wijmenga, "Shared genetics in coeliac disease and other immune-mediated diseases," *Journal of Internal Medicine*, vol. 269, no. 6, pp. 591–603, 2011.
- [4] J. M. Anaya, Y. Shoenfeld, P. A. Correa, M. García-Carrasco, and R. Cervera, *Autoinmunidad y Enfermedad Autoinmune*, Corporación para Investigaciones Biológicas, Medellín, Colombia, 1st edition, 2005.
- [5] C. Cotsapas, B. F. Voight, E. Rossin et al., "Pervasive sharing of genetic effects in autoimmune disease," *PLoS Genetics*, vol. 7, no. 8, Article ID e1002254, 2011.
- [6] A. Delgado-Vega, E. Sánchez, S. Löfgren, C. Castillejo-López, and M. E. Alarcón-Riquelme, "Recent findings on genetics of systemic autoimmune diseases," *Current Opinion in Immunology*, vol. 22, no. 6, pp. 698–705, 2010.
- [7] M. Sirota, M. A. Schaub, S. Batzoglou, W. H. Robinson, and A. J. Butte, "Autoimmune disease classification by inverse association with SNP alleles," *PLoS Genetics*, vol. 5, no. 12, Article ID e1000792, 2009.
- [8] M. Weeber, J. A. Kors, and B. Mons, "Online tools to support literature-based discovery in the life sciences," *Briefings in Bioinformatics*, vol. 6, no. 3, pp. 277–286, 2005.
- [9] Y. Tsuruoka, M. Miwa, K. Hamamoto, J. Tsujii, and S. Ananiadou, "Discovering and visualizing indirect associations between biomedical concepts," *Bioinformatics*, vol. 27, no. 13, pp. i111–i119, 2011.
- [10] E. Tjioe, M. W. Berry, and R. Homayouni, "Discovering gene functional relationships using FAUN (Feature Annotation Using Nonnegative matrix factorization)," *BMC Bioinformatics*, vol. 11, no. 6, article 14, 2010.
- [11] R. Rodriguez-Esteban, "Biomedical text mining and its applications," *PLoS Computational Biology*, vol. 5, no. 12, Article ID e1000597, 2009.
- [12] I. Spasic, S. Ananiadou, J. McNaught, and A. Kumar, "Text mining and ontologies in biomedicine: making sense of raw text," *Briefings in Bioinformatics*, vol. 6, no. 3, pp. 239–251, 2005.
- [13] R. Jelier, M. J. Schuemie, P. J. Roes, E. M. van Mulligen, and J. A. Kors, "Literature-based concept profiles for gene annotation: the issue of weighting," *International Journal of Medical Informatics*, vol. 77, no. 5, pp. 354–362, 2008.
- [14] M. S. Siadat and W. A. Knaus, "Locating previously unknown patterns in data-mining results: a dual data- and knowledge-mining method," *BMC Medical Informatics and Decision Making*, vol. 6, article 13, 2006.
- [15] A. Ma'ayan, R. D. Blitzer, and R. Iyengar, "Toward predictive models of mammalian cells," *Annual Review of Biophysics and Biomolecular Structure*, vol. 34, pp. 319–349, 2005.
- [16] T. Aittokallio and B. Schwikowski, "Graph-based methods for analysing networks in cell biology," *Briefings in Bioinformatics*, vol. 7, no. 3, pp. 243–255, 2006.
- [17] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, and R. Guthke, "Gene regulatory network inference: data integration in dynamic models-A review," *BioSystems*, vol. 96, no. 1, pp. 86–103, 2009.
- [18] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [19] E. G. Cerami, G. D. Bader, B. E. Gross, and C. Sander, "cPath: open source software for collecting, storing, and querying biological pathways," *BMC Bioinformatics*, vol. 7, article 497, 2006.
- [20] H. Hermjakob, L. Montecchi-Palazzi, G. Bader et al., "The HUPO PSI's molecular Interaction format—a community standard for the representation of protein interaction data," *Nature Biotechnology*, vol. 22, no. 2, pp. 177–183, 2004.
- [21] J. Garcia-Garcia, E. Guney, R. Aragues, J. Planas-Iglesias, and B. Oliva, "Biana: a software framework for compiling biological interactions and analyzing networks," *BMC Bioinformatics*, vol. 11, article 56, 2010.
- [22] A. Nikitin, S. Egorov, N. Daraselia, and I. Mazo, "Pathway studio—the analysis and navigation of molecular networks," *Bioinformatics*, vol. 19, no. 16, pp. 2155–2157, 2003.
- [23] A. Yuryev, Z. Mulyukov, E. Kotelnikova et al., "Automatic pathway building in biological association networks," *BMC Bioinformatics*, vol. 7, article 171, 2006.
- [24] U. Dogrusoz, A. Cetintas, E. Demir, and O. Babur, "Algorithms for effective querying of compound graph-based pathway databases," *BMC Bioinformatics*, vol. 10, article 376, 2009.
- [25] J. Chen, B. J. Aronow, and A. G. Jegga, "Disease candidate gene identification and prioritization using protein interaction networks," *BMC Bioinformatics*, vol. 10, article 73, 2009.
- [26] X. Zhu, M. Gerstein, and M. Snyder, "Getting connected: analysis and principles of biological networks," *Genes and Development*, vol. 21, no. 9, pp. 1010–1024, 2007.
- [27] R. Jelier, G. Jenster, L. C. J. Dorssers et al., "Text-derived concept profiles support assessment of DNA microarray data for acute myeloid leukemia and for androgen receptor stimulation," *BMC Bioinformatics*, vol. 8, article 14, 2007.
- [28] R. Jelier, M. J. Schuemie, A. Veldhoven, L. C. J. Dorssers, G. Jenster, and J. A. Kors, "Anni 2.0: a multipurpose text-mining tool for the life sciences," *Genome Biology*, vol. 9, no. 6, article R96, 2008.
- [29] S. I. Berger, J. M. Posner, and A. Ma'ayan, "Genes2Networks: connecting lists of gene symbols using mammalian protein interactions databases," *BMC Bioinformatics*, vol. 8, article 372, 2007.
- [30] R. A. Cifuentes, A. Rojas-Villarraga, and J. M. Anaya, "Human leukocyte antigen class II and type 1 diabetes in Latin America: a combined meta-analysis of association and family-based studies," *Human Immunology*, vol. 72, no. 7, pp. 581–586, 2011.
- [31] A. E. Handel, L. Handunnetthi, A. J. Berlanga, C. T. Watson, J. M. Morahan, and S. V. Ramagopalan, "The effect of single nucleotide polymorphisms from genome wide association studies in multiple sclerosis on gene expression," *PLoS ONE*, vol. 5, no. 4, Article ID e10142, 2010.
- [32] D. Plant, E. Flynn, H. Mbarek et al., "Investigation of potential non-HLA rheumatoid arthritis susceptibility loci in a European cohort increases the evidence for nine markers," *Annals of the Rheumatic Diseases*, vol. 69, no. 8, pp. 1548–1553, 2010.
- [33] P. K. Gregersen, C. I. Amos, A. T. Lee et al., "REL, encoding a member of the NF- κ B family of transcription factors, is a newly

- defined risk locus for rheumatoid arthritis," *Nature Genetics*, vol. 41, no. 7, pp. 820–823, 2009.
- [34] V. Gateva, J. K. Sandling, G. Hom et al., "A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus," *Nature Genetics*, vol. 41, no. 11, pp. 1228–1233, 2009.
- [35] L. Zhang, W. Li, L. Song, and L. Chen, "A towards-multidimensional screening approach to predict candidate genes of rheumatoid arthritis based on SNP, structural and functional annotations," *BMC Medical Genomics*, vol. 3, article 38, 2010.
- [36] Y. M. Hussein, S. A. El Tarhouny, R. H. Mohamed, A. S. El-Shal, A. M. Abul-Saoud, and M. Abdo, "Association of CD4 enhancer gene polymorphisms with rheumatoid arthritis in Egyptian female patients," *Rheumatology International*. In press.
- [37] S. F. Lo, L. Wan, H. C. Lin, C. M. Huang, and F. J. Tsai, "Association of CD4 enhancer gene polymorphisms with rheumatoid arthritis and systemic lupus erythematosus in Taiwan," *Journal of Rheumatology*, vol. 35, no. 11, pp. 2113–2118, 2008.
- [38] J. Schulte-Pelkum, M. Fritzler, and M. Mahler, "Latest update on the Ro/SS-A autoantibody system," *Autoimmunity Reviews*, vol. 8, no. 7, pp. 632–637, 2009.
- [39] M. Dugar, S. Cox, V. Limaye, T. P. Gordon, and P. J. Roberts-Thomson, "Diagnostic utility of anti-Ro52 detection in systemic autoimmunity," *Postgraduate Medical Journal*, vol. 86, no. 1012, pp. 79–82, 2010.
- [40] M. Tanaka, K. Tanji, M. Niida, and T. Kamitani, "Dynamic movements of Ro52 cytoplasmic bodies along microtubules," *Histochemistry and Cell Biology*, vol. 133, no. 3, pp. 273–284, 2010.
- [41] A. Espinosa, V. Dardalhon, S. Brauner et al., "Loss of the lupus autoantigen Ro52/Trim21 induces tissue inflammation and systemic autoimmunity by dysregulating the IL-23-Th17 pathway," *Journal of Experimental Medicine*, vol. 206, no. 8, pp. 1661–1671, 2009.
- [42] A. P. M. Verhagen and G. J. M. Pruijn, "Are the Ro RNP-associated Y RNAs concealing microRNAs? Y RNA-derived miRNAs may be involved in autoimmunity," *BioEssays*, vol. 33, no. 9, pp. 674–682, 2011.
- [43] A. Reich, M. Meurer, A. Viehweg, and D. J. Muller, "Narrow-band UVB-induced externalization of selected nuclear antigens in keratinocytes: implications for lupus erythematosus pathogenesis," *Photochemistry and Photobiology*, vol. 85, no. 1, pp. 1–7, 2009.
- [44] H. Park, L. Yu, T. Kim, B. Cho, J. Kang, and Y. Park, "Antigenic determinants to GAD autoantibodies in patients with type 1 diabetes with and without autoimmune thyroid disease," *Annals of the New York Academy of Sciences*, vol. 1079, pp. 213–219, 2006.
- [45] A. Rojas-Villarraga, D. Botello-Corzo, and J. M. Anaya, "HLA-Class II in Latin American patients with type 1 diabetes," *Autoimmunity Reviews*, vol. 9, no. 10, pp. 666–673, 2010.
- [46] N. Castaño-Rodríguez, L. M. Diaz-Gallo, R. Pineda-Tamayo, A. Rojas-Villarraga, and J. M. Anaya, "Meta-analysis of HLA-DRB1 and HLA-DQB1 polymorphisms in Latin American patients with systemic lupus erythematosus," *Autoimmunity Reviews*, vol. 7, no. 4, pp. 322–330, 2008.
- [47] O. L. Rojas, A. Rojas-Villarraga, P. Cruz-Tapias et al., "HLA class II polymorphism in Latin American patients with multiple sclerosis," *Autoimmunity Reviews*, vol. 9, no. 6, pp. 407–413, 2010.
- [48] T. Kamradt and N. Avrión Mitchison, "Tolerance and autoimmunity," *New England Journal of Medicine*, vol. 344, no. 9, pp. 655–664, 2001.
- [49] A. Korb, H. Pavenstädt, and T. Pap, "Cell death in rheumatoid arthritis," *Apoptosis*, vol. 14, no. 4, pp. 447–454, 2009.
- [50] A. Pawlik, M. Herczyńska, M. Kurzawski et al., "IL-1 β , IL-6 and TNF gene polymorphisms do not affect the treatment outcome of rheumatoid arthritis patients with leflunomide," *Pharmacological Reports*, vol. 61, no. 2, pp. 281–287, 2009.
- [51] I. N. Sarkar, "Biomedical informatics and translational medicine," *Journal of Translational Medicine*, vol. 8, article 22, 2010.

