



# Calculation of Similarity Between 26 Autoimmune Diseases Based on Three Measurements Including Network, Function, and Semantics

Yanjun Ding<sup>1,2</sup>, Mintian Cui<sup>1</sup>, Jun Qian<sup>2</sup>, Chao Wang<sup>1</sup>, Qi Shen<sup>1</sup>, Hongbiao Ren<sup>1</sup>, Liangshuang Li<sup>1</sup>, Fengmin Zhang<sup>2\*</sup> and Ruijie Zhang<sup>1\*</sup>

<sup>1</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, <sup>2</sup>Department of Microbiology, WU Lien-Teh Institute, Harbin Medical University, Harbin, China

## OPEN ACCESS

### Edited by:

Madhuchanda Bhattacharjee,  
University of Hyderabad, India

### Reviewed by:

Juan-Manuel Anaya,  
Rosario University, Colombia  
Tanima Bose,  
Ludwig Maximilian University of  
Munich, Germany  
Carolina Ramirez-Santana,  
Rosario University, Colombia

### \*Correspondence:

Ruijie Zhang  
zhangrjhm@163.com  
Fengmin Zhang  
fengminzhang@ems.hrbmu.edu.cn

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 13 August 2021

Accepted: 27 October 2021

Published: 11 November 2021

### Citation:

Ding Y, Cui M, Qian J, Wang C,  
Shen Q, Ren H, Li L, Zhang F and  
Zhang R (2021) Calculation of Similarity  
Between 26 Autoimmune Diseases  
Based on Three Measurements  
Including Network, Function,  
and Semantics.  
Front. Genet. 12:758041.  
doi: 10.3389/fgene.2021.758041

Autoimmune diseases (ADs) are a broad range of diseases in which the immune response to self-antigens causes damage or disorder of tissues, and the genetic susceptibility is regarded as the key etiology of ADs. Accumulating evidence has suggested that there are certain commonalities among different ADs. However, the theoretical research about similarity between ADs is still limited. In this work, we first computed the genetic similarity between 26 ADs based on three measurements: network similarity (NetSim), functional similarity (FunSim), and semantic similarity (SemSim), and systematically identified three significant pairs of similar ADs: rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), myasthenia gravis (MG) and autoimmune thyroiditis (AIT), and autoimmune polyendocrinopathies (AP) and uveomeningoencephalitic syndrome (Vogt-Koyanagi-Harada syndrome, VKH). Then we investigated the gene ontology terms and pathways enriched by the three significant AD pairs through functional analysis. By the cluster analysis on the similarity matrix of 26 ADs, we embedded the three significant AD pairs in three different disease clusters respectively, and the ADs of each disease cluster might have high genetic similarity. We also detected the risk genes in common among the ADs which belonged to the same disease cluster. Overall, our findings will provide significant insight in the commonalities of different ADs in genetics, and contribute to the discovery of novel biomarkers and the development of new therapeutic methods for ADs.

**Keywords:** ADs, genetic susceptibility, network similarity, functional similarity, semantic similarity, autoimmune tautology

## INTRODUCTION

Autoimmune diseases (ADs) are characterized by causing abnormal immune response which can damage human tissues as a result of the loss of immune tolerance to self-antigens (Margo and Harman, 2016). ADs affect more than 5% of the global population, the incidence and mortality of which have also increased markedly (Ji et al., 2016). Possible causes contain genetic, environmental, hormonal, and immunological factors (Stojanovich and Marisavljevich, 2008). However, neither the inner mechanism action nor the etiology of ADs is clear and there is still no effective cure for these diseases (Rosenblum et al., 2012; Li et al., 2017).

ADs share several clinical signs and symptoms, physiopathological mechanisms, and environmental and genetic factors, and this fact indicates that they have a common origin, which has been called the autoimmune tautology. A growing body of evidence has indicated the existence of the autoimmune tautology among various ADs: 1) Different ADs exhibit the same phenotypic characteristics (Anaya, 2017). These diseases, whether organ-specific or systematic, show tissue and organ damage and inflammatory pathological features (Place and Kanneganti, 2020). 2) Different ADs exhibit the same clinical characteristics. Clinically, the results from serological examinations of patients often overlap. And the same patient may suffer from two or more ADs simultaneously, which has been called the polyautoimmunity (PolyA) (Anaya, 2014). In addition, there is a tendency for ADs to cluster within families (Cardenas-Roldan et al., 2013). 3) Different ADs exhibit the same genetic characteristics. ADs are caused by the mutation of multiple loci in the human genome and share the same main genetic loci (Anaya et al., 2006). For example, a previous study indicated that the human leukocyte antigen (HLA) is a susceptibility gene shared by multiple ADs (Cruz-Tapias et al., 2012). And Ueda et al. (2003) found that a molecule encoded by *CTLA4* was vital for negative regulation of the immune system and could enhance the risk of several ADs, such as Graves disease, autoimmune hypothyroidism, and type 1 diabetes mellitus (T1D), which indicated that ADs might share similar pathogenic mechanisms. Li et al. (2015) proved that different pediatric ADs shared the same genetic variation. They analyzed the clinical cases of ten different ADs and found many of these diseases were familial and the patients often suffered from several ADs at the same time. In this study, 27 significant risk genetic loci were identified, of which 22 were shared by at least two ADs and 19 loci were shared by at least three ADs. Thus, identification of risk genes shared by multiple ADs may help to explain the development of PolyA. 4) In addition, different ADs also exhibit the same epigenetic characteristics. Epigenetic researches found that ADs shared similar epigenetic mechanisms. For instance, the DNA promoter region in the target cells of systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) showed low methylation (Quintero-Ronderos and Montoya-Ortiz, 2012). The strong similarity among ADs provides us with a deeper understanding of the common underlying mechanisms of ADs, and also prompts researchers to classify ADs. Therefore, the studies on the genetic similarity of ADs can help us to dissect AD pathogenesis, and contribute to the discovery of novel biomarkers and the development of new therapeutic methods for ADs, which is extremely important in clinical research.

In this study, we utilized three measurements, including network similarity, functional similarity, and semantic similarity, to analyze genetic similarity between 26 ADs (the workflow diagram is shown in **Figure 1**). We identified three significant pairs of similar ADs by multi-step computational approaches. Besides, based on the similarity matrix of 26 ADs, we found some other ADs which were similar to significant pairs of similar ADs by cluster analysis. And the risk genes shared by the ADs which belonged to the same disease cluster

could be promising biomarkers for ADs. Our findings provided a novel perspective to understand the commonalities of different ADs in genetics and would facilitate AD mechanism research.

## MATERIALS AND METHODS

### Collection of AD Terms and AD-Related Genes

The AD terms (category C20.111), including 68 diseases, were acquired from the Medical Subject Headings (MeSH, <https://www.nlm.nih.gov/mesh/meshhome.html>). After removing the complications of ADs, we extracted human disease-related genes from the Genetic Association Database (GAD, <https://geneticassociationdb.nih.gov/>) (Becker et al., 2004) and mapped these genes to the AD terms for integration. The disease gene sets consisted of 267 related genes of 26 ADs (**Supplementary Table S1**).

### Calculation of Network Similarity Between ADs

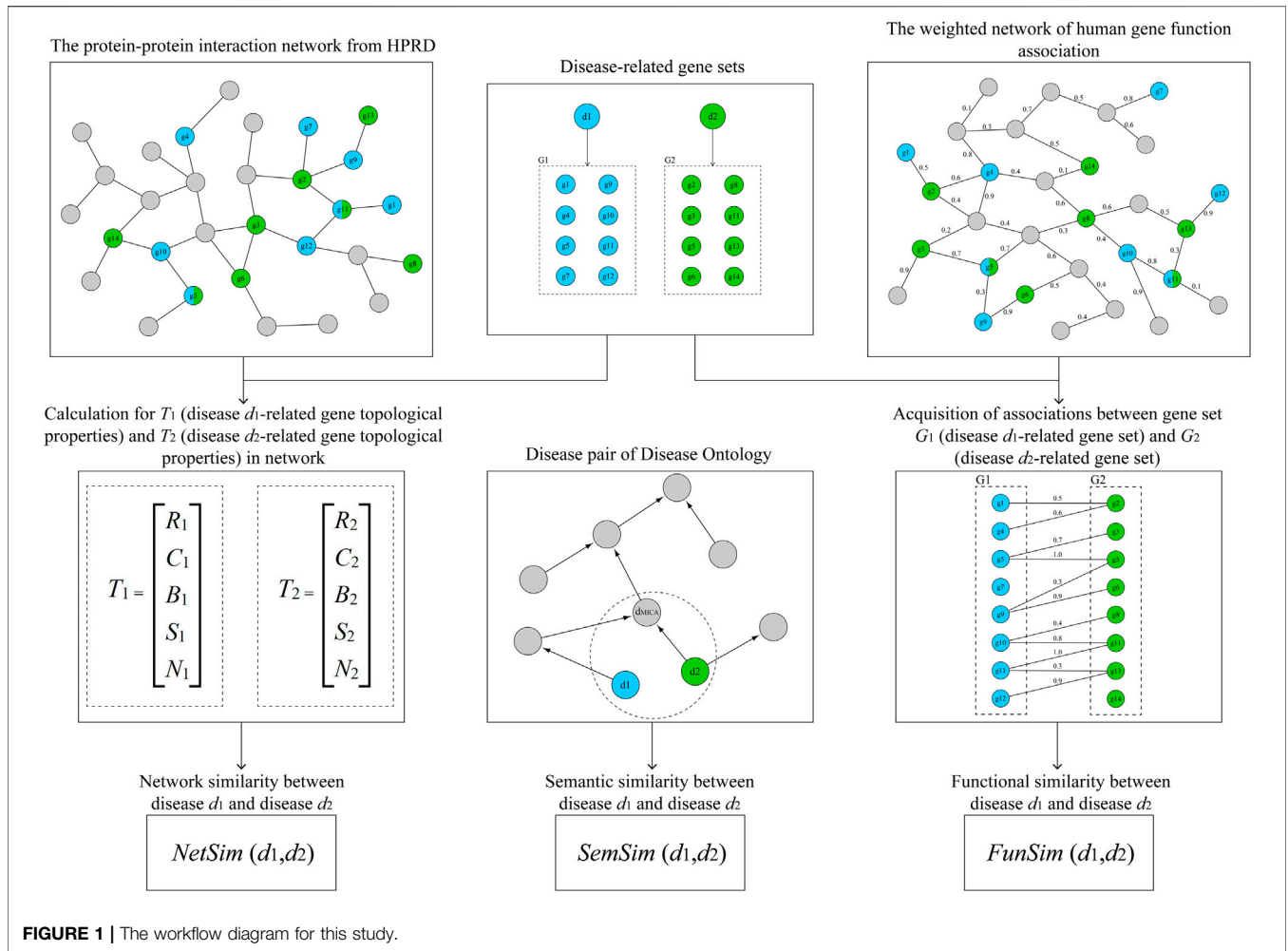
We downloaded the information on protein-protein interactions of human genes from the Human Protein Reference Database (HPRD, <http://www.hprd.org/>) (Peri et al., 2003) and used Cytoscape software (v3.8.2) (Shannon et al., 2003) to construct a human protein-protein interaction network. The topological properties of AD-related genes in this network were computed (Hidalgo et al., 2009; Chavali et al., 2010). The gene set of disease  $d$  was defined as  $G = \{g_1, g_2, g_3, \dots, g_i, \dots, g_k\}$ . In order to assess the network similarity between ADs, we first calculated the average topological properties of each AD in the network as follows:

$$R = \frac{\sum_{1 \leq i \leq k} r_i}{k}, C = \frac{\sum_{1 \leq i \leq k} c_i}{k}, B = \frac{\sum_{1 \leq i \leq k} b_i}{k}, S = \frac{\sum_{1 \leq i \leq k} s_i}{k}, \text{ and} \\ H = \frac{\sum_{1 \leq i \leq k} h_i}{k} \quad (1-5)$$

where  $k$  represents the number of genes in  $G$ ,  $g_i$  is the  $i$ th gene of  $G$ ,  $r_i$  is the degree of  $g_i$ ,  $c_i$  is the clustering coefficient of  $g_i$ ,  $b_i$  is the betweenness centrality of  $g_i$ ,  $s_i$  is the average shortest path length of  $g_i$ ,  $h_i$  is the neighborhood connectivity of  $g_i$ ,  $R$  is the degree of  $d$ ,  $C$  is the clustering coefficient of  $d$ ,  $B$  is the betweenness centrality of  $d$ ,  $S$  is the average shortest path length of  $d$ , and  $H$  is the neighborhood connectivity of  $d$ . As shown in **Figure 1**,  $d_1$  and  $d_2$  are two ADs from MeSH, and  $G_1$  and  $G_2$  are gene sets related to  $d_1$  and  $d_2$ . The average topological properties of  $d_1$  and  $d_2$  were defined as vector  $T_1 = \{R_1, C_1, B_1, S_1, H_1\}$  and vector  $T_2 = \{R_2, C_2, B_2, S_2, H_2\}$ , respectively. We defined the network similarity (NetSim) score between  $d_1$  and  $d_2$  as the Pearson correlation coefficient (PCC) calculated with  $T_1$  and  $T_2$ . The formula that was used as follows:

$$\rho_{T_1, T_2} = \frac{\text{cov}(T_1, T_2)}{\sigma_{T_1} \sigma_{T_2}} \quad (6)$$

where  $\text{cov}(T_1, T_2)$  is the covariance of variables  $T_1$  and  $T_2$ ,  $\sigma_{T_1}$  and  $\sigma_{T_2}$  are the standard deviations for  $T_1$  and  $T_2$ .



### Calculation of Functional Similarity Between ADs

The data on the functional interactions of genes was downloaded from HumanNet, which is a human gene functional interaction network based on Gene Ontology annotation (Lee et al., 2011). Each interaction in HumanNet has a log likelihood score (LLS) that measures the probability of a functional association between genes (Cheng et al., 2014).

We downloaded LLSs between human genes from HumanNet and normalized the LLSs as follows:

$$LLS_N(g_i, g_j) = \frac{LLS(g_i, g_j) - LLS_{min}}{LLS_{max} - LLS_{min}} \tag{7}$$

where  $g_i$  and  $g_j$  are the  $i$ th and  $j$ th gene respectively.  $LLS_N(g_i, g_j)$  indicates LLS between  $g_i$  and  $g_j$  after normalization.  $LLS(g_i, g_j)$  indicates LLS between  $g_i$  and  $g_j$ .  $LLS_{min}$  and  $LLS_{max}$  are the minimum LLS and the maximum LLS of HumanNet respectively.

The functional similarity (FunSim) score between a pair of genes was defined as follows:

$$FunSim(g_i, g_j) = \begin{cases} 1 & i = j \\ LLS_N & i \neq j \text{ and } e(i, j) \in E(HumanNet) \\ 0 & i \neq j \text{ and } e(i, j) \notin E(HumanNet) \end{cases} \tag{8}$$

where  $e(i, j)$  represents the interaction edge between  $g_i$  and  $g_j$ .  $E(HumanNet)$  is a set including all the edges of HumanNet.

Next, we defined the functional association between a gene  $g$  and a gene set  $G = \{g_1, g_2, g_3, \dots, g_i, \dots, g_k\}$  as follows:

$$F_G(g) = \max_{1 \leq i \leq k} (FunSim(g, g_i)), \quad g_i \in G \tag{9}$$

where  $k$  represents the number of genes in  $G$ ,  $g_i$  is the  $i$ th gene of  $G$ .

$G_1 = \{g_{11}, g_{12}, \dots, g_{1i}, \dots, g_{1m}\}$  and  $G_2 = \{g_{21}, g_{22}, \dots, g_{2j}, \dots, g_{2n}\}$  are gene sets related to  $d_1$  and  $d_2$  respectively.  $m$  is the number of genes in  $G_1$ , and  $n$  is the number of genes in  $G_2$ . We defined FunSim score of  $d_1$  and  $d_2$  as follows:

$$FunSim(d_1, d_2) = \frac{\sum_{1 \leq i \leq m} F_{G_2}(g_{1i}) + \sum_{1 \leq j \leq n} F_{G_1}(g_{2j})}{m + n}, \quad g_{1i} \in G_1, \quad g_{2j} \in G_2 \tag{10}$$

## Calculation of Semantic Similarity Between ADs

Semantic similarity is a method to measure the closeness between two terms, according to a given ontology (Zhang and Lai, 2015; Zhang and Lai, 2016; Del Prete et al., 2018). The Resnik method was applied to our study. The human disease terms were obtained from the Human Disease Ontology (DO, <http://www.disease-ontology.org>) (Schriml et al., 2019). The DO includes the breadth of common and rare diseases, organized as a directed acyclic graph in which a term represents a DO term and an edge represents an “IS\_A” relationship between diseases. The information content (IC) of each DO term could be calculated as follows:

$$IC(d) = -\log\left(\frac{n}{N}\right) \quad (11)$$

where  $d$  is a disease term of DO,  $n$  is the number of genes related to  $d$ , and  $N$  is the total number of genes related to DO. As shown in **Figure 1**,  $d_1$  and  $d_2$  are two AD terms of DO, and  $d_{MICA}$  is the most informative common ancestor (MICA) of  $d_1$  and  $d_2$ . The MICA means the ancestor that has the maximum IC among all the common ancestors between terms of ontology. And we defined the semantic similarity (SemSim) score of  $d_1$  and  $d_2$  as follows:

$$SemSim(d_1, d_2) = IC(d_{MICA}) \quad (12)$$

## Calculation of Integrated Similarity Between ADs

To identify more reliable pairs of similar ADs, we integrated above three kinds of similarity (NetSim, FunSim, and SemSim) scores to comprehensively determine the levels of similarity between ADs as follows:

$$IntegratedSim(d_1, d_2) = \sqrt[3]{NetSim(d_1, d_2) \cdot FunSim(d_1, d_2) \cdot SemSim(d_1, d_2)} \quad (13)$$

where  $d_1$  and  $d_2$  are two AD terms of MeSH.

## Functional Analysis of Related Genes of Similar ADs

Functional enrichment analysis of Gene Ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) for related genes of similar ADs was performed to infer potential biological processes and pathways using the DAVID Bioinformatics Tool (<http://david.abcc.ncifcrf.gov/>, version 6.7) (Huang Da et al., 2009). The  $p$ -values for the biological processes and pathways were adjusted for false discovery rate (FDR) by the Benjamini-Hochberg method. The biological processes and pathways with FDR less than 0.05 were considered statistically significant functional categories.

## Cluster Analysis for ADs Based on Three Kinds of Similarity

To determine whether multiple ADs had high genetic similarity, hierarchical clustering was performed on the integrated similarity scores between 26 ADs based on the Euclidean distance. The  $d_1$

and  $d_2$  are two diseases of 26 ADs. We defined the Euclidean distance between  $d_1$  and  $d_2$  as follows:

$$E(d_1, d_2) = \sqrt{\sum_{1 \leq i \leq 26} (x_{1i} - x_{2i})^2} \quad (14)$$

where  $x_{1i}$  is the integrated similarity score between  $d_1$  and  $i$ th disease of 26 ADs,  $x_{2i}$  is the integrated similarity score between  $d_2$  and  $i$ th disease of 26 ADs. The 26 ADs were clustered based on the Euclidean distances between diseases and there were shorter Euclidean distances between ADs belonging to the same disease cluster.

## RESULTS

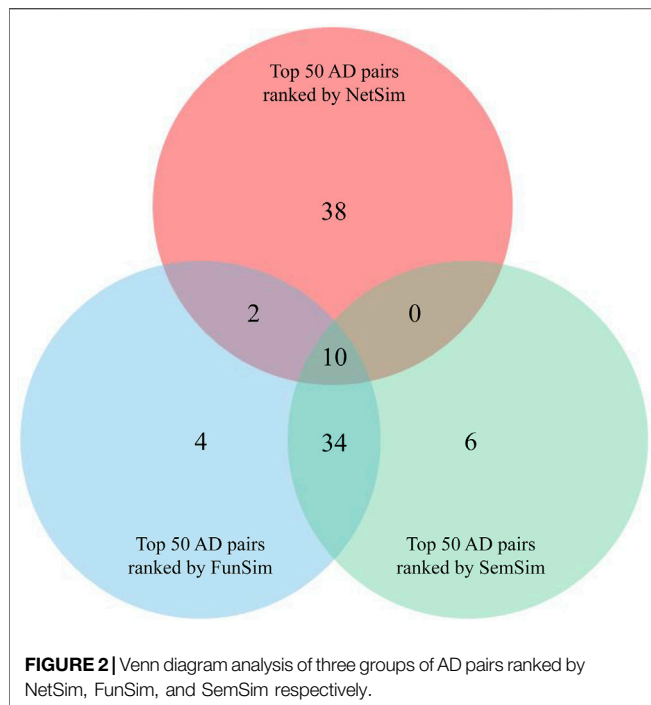
### Identification of Potential Pairs of Similar ADs

To identify pairs of similar diseases in genetics, we calculated similarity between 26 ADs based on three measurements. All of AD pairs were sorted in descending according to their scores of NetSim, FunSim, and SemSim. To enhance the reliability of the study, we selected ten AD pairs from intersection of top 50 AD pairs with the highest NetSim score, top 50 AD pairs with the highest FunSim score, and top 50 AD pairs with the highest SemSim score for further analysis (**Figure 2** and **Supplementary Tables S2–4**), which were considered as potential pairs of similar ADs.

Further, a network was generated on the ten AD pairs and their related genes, including 14 ADs and 247 genes (**Figure 3**). The topological properties of the network were investigated. We extracted 35 genes each of whose degree was greater than or equal to three from the network as the possible AD relevant genes. The top five genes regarding degree were *HLA-DRB1*, *HLA-DQB1*, *CTLA4*, *HLA-DQA1*, and *TNF*, indicating that these genes were critical in multiple ADs. Previous researches have demonstrated that these genes are associated with many ADs and exert various functions in human autoimmune disorders. Besides, we ascertained that RA and multiple sclerosis (MS) involved more similarity relationships than other AD in this network, implying that the pathogenesis of RA and MS might exist in most of ADs from the network. What's more, a pair of diseases with higher number of shared genes, suggesting they likely have higher genetic similarity. Thus, the ten potential pairs of similar ADs were ranked by their number of shared genes and shown in **Table 1**. And the top-rank disease pair is RA and SLE, followed by T1D and RA, and RA and MS.

### Functional Implication of the Genes Related to Multiple ADs

To explore common genetic mechanisms of a variety of ADs, we performed functional enrichment analysis of GO and KEGG for the 35 AD relevant genes (**Figure 4A**). The cutoff criterion was a FDR less than 0.05. The top ten significant GO terms in the BP were mainly associated with immune response, antigen processing and presentation, and interferon gamma (IFN $\gamma$ )-related functions (**Figure 4B**). Notably, *IFNG* was involved in



the top three GO terms and was defined as a hub gene. *IFNG* can encode IFN $\gamma$  that is a cytokine that is critical for innate and adaptive immunity against viral, bacterial, and protozoan infections. And aberrant IFN $\gamma$  expression is associated with a number of ADs, such as RA and SLE (Hu and Ivashkiv, 2009; Barrat et al., 2019). The top ten significant KEGG pathways contained four AD-correlated pathways, such as “inflammatory bowel disease (IBD),” “autoimmune thyroid disease,” “type 1 diabetes mellitus,” and “rheumatoid arthritis,” which illustrated that the 35 genes might induce the initiation and development of multiple ADs (Figure 4C). *HLA-DQB1*, *HLA-DRB1*, *HLA-DPBI*, *HLA-DQA1* were involved in all of the top ten KEGG pathways and were defined as hub genes. The four genes belonged to HLA class II alleles which were suggested to contribute to the susceptibility and resistance to ADs (Wieber et al., 2021).

### Identification of Significant Pairs of Similar ADs

To identify more reliable pairs of similar ADs, we integrated the three measurements to compute integrated similarity scores between 26 ADs (see Materials and Methods). All of AD pairs were sorted in descending according to their integrated similarity scores. And the top ten AD pairs were extracted for further analysis (Table 2). We found that the ten AD pairs contained three potential pairs of similar ADs consisting of RA and SLE, myasthenia gravis (MG) and autoimmune thyroiditis (AIT), and autoimmune polyendocrinopathies (AP) and uveomeningoencephalitic syndrome (Vogt-Koyanagi-Harada syndrome, VKH) which were defined as the significant pairs of similar ADs.

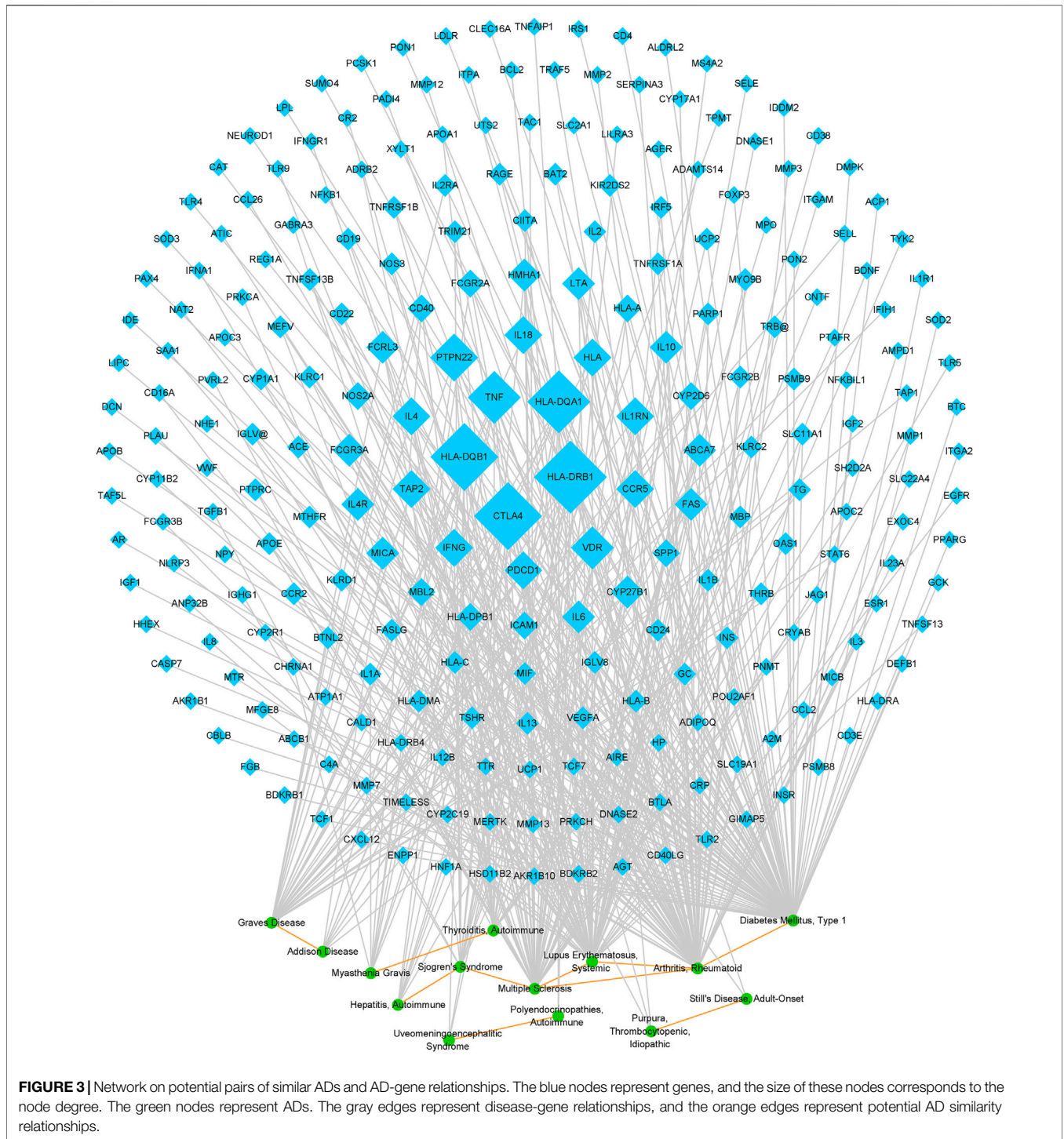
### Functional Analysis of Related Genes of Significant Pairs of Similar ADs

To reveal the underlying mechanisms shared by two similar ADs, we performed functional enrichment analysis of related genes of RA and SLE, MG and AIT, and AP and VKH. The cutoff criterion was a FDR less than 0.05. The related genes of RA and SLE were significantly enriched in GO terms mainly involved in immune response, inflammatory response, and IFN $\gamma$ . The significant enriched pathways including RA, inflammatory bowel disease (IBD), tuberculosis, etc (Figures 5A,B). The related genes of MG and AIT were mainly related to immune response (GO), antigen processing and presentation (GO), autoimmune thyroid disease (AITD) (KEGG), and allograft rejection (KEGG) (Figures 5C,D). Moreover, the related genes of AP and VKH were mainly associated with the antigen processing and presentation in GO and some pathways such as viral myocarditis, *Staphylococcus aureus* infection, AITD, intestinal immune network for IgA production, etc (Figures 5E,F).

### Hierarchical Clustering Result of 26 ADs Based on Three Similarity Measurements

To determine whether there was high genetic similarity among multiple ADs, we applied hierarchical clustering to the integrated similarity matrix of 26 ADs. The disease clusters consisting of AD pairs with integrated similarity scores greater than 0.3 were considered to be significant. As shown in Figure 6A, three significant disease groups were identified from the 26 ADs, and the ADs of each disease group might have high genetic similarity. We found that the three significant pairs of similar ADs were located in three different clusters, respectively. In the cluster one, bullous pemphigoid might be similar to AP and VKH. The ADs of cluster one are involved in endocrine autoimmunity. For instance, bullous pemphigoid has been proved to be related to immunodysregulation polyendocrinopathy enteropathy X-linked syndrome (Mcginness et al., 2006). In the cluster two, pemphigus, Graves disease, Sjogren’s syndrome, Addison Disease, and autoimmune hepatitis might be similar to MG and AIT. The ADs of cluster two contain the main AITDs (AIT and Graves disease) and frequent ADs involved in PolyA (AIT, Graves disease, and Sjogren’s syndrome) (Amador-Patarroyo et al., 2012; Botello et al., 2020). In the cluster three, T1D and MS might be similar to RA and SLE. The ADs of cluster three are all chronic inflammatory ADs and share multiple genetic susceptibility loci (Richard-Miceli and Criswell, 2012). Then, we performed pathway enrichment analysis of the related genes of ADs of three significant clusters. The related genes of ADs of three disease clusters were significantly enriched in pathways of T1D and IBD (Figures 6B–D). Therefore, we infer that T1D and IBD can participate in PolyA in various AD patients.

Next, we detected the causal genes in common among the ADs which belonged to the same significant disease cluster, and these



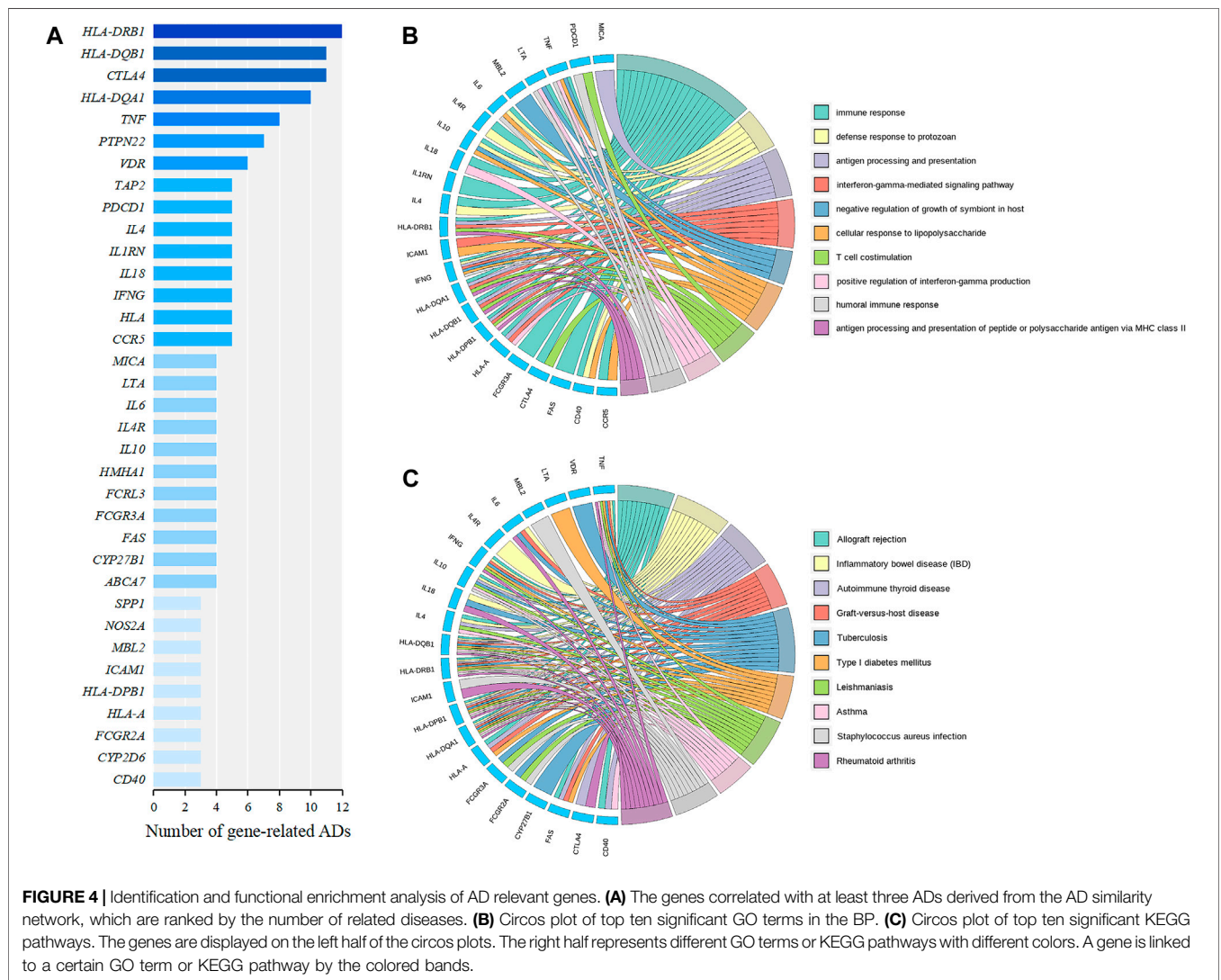
genes could be used for PolyA research. As shown in **Table 3**, the ADs of cluster one shared one gene (*HLA-DQA1*); the ADs of cluster two shared two genes (*HLA-DQB1* and *HLA-DRB1*); the ADs of cluster three shared eight genes (*TNF*, *HLA-DRB1*, *PDCD1*, *PTPN22*, *CCR5*, *IL6*, *HLA-DQB1*, and *CTLA4*). Identification of these genes will contribute to the discovery of novel prognostic, diagnostic, and therapeutic markers and justification of drug repurposing for ADs.

## DISCUSSION

During the past years, numerous studies have confirmed that different ADs are similar in various aspects. Nevertheless, these studies just focused on several ADs, and lacking a comprehensive analysis on similarity between ADs from the perspective of genetics. To date, various disease similarity methods have been developed (Dozmorov, 2019).

**TABLE 1 |** The ten potential pairs of similar ADs ranked by number of shared genes.

Rank	Autoimmune disease	Autoimmune disease	Number of shared gene
1	Arthritis, Rheumatoid	Lupus Erythematosus, Systemic	40
2	Diabetes Mellitus, Type 1	Arthritis, Rheumatoid	28
3	Arthritis, Rheumatoid	Multiple Sclerosis	23
4	Lupus Erythematosus, Systemic	Multiple Sclerosis	20
5	Multiple Sclerosis	Sjogren's Syndrome	9
6	Myasthenia Gravis	Thyroiditis, Autoimmune	7
7	Addison Disease	Graves Disease	6
8	Hepatitis, Autoimmune	Sjogren's Syndrome	3
9	Polyendocrinopathies, Autoimmune	Uveomeningoencephalitic Syndrome	2
10	Purpura, Thrombocytopenic, Idiopathic	Still's Disease, Adult-Onset	1

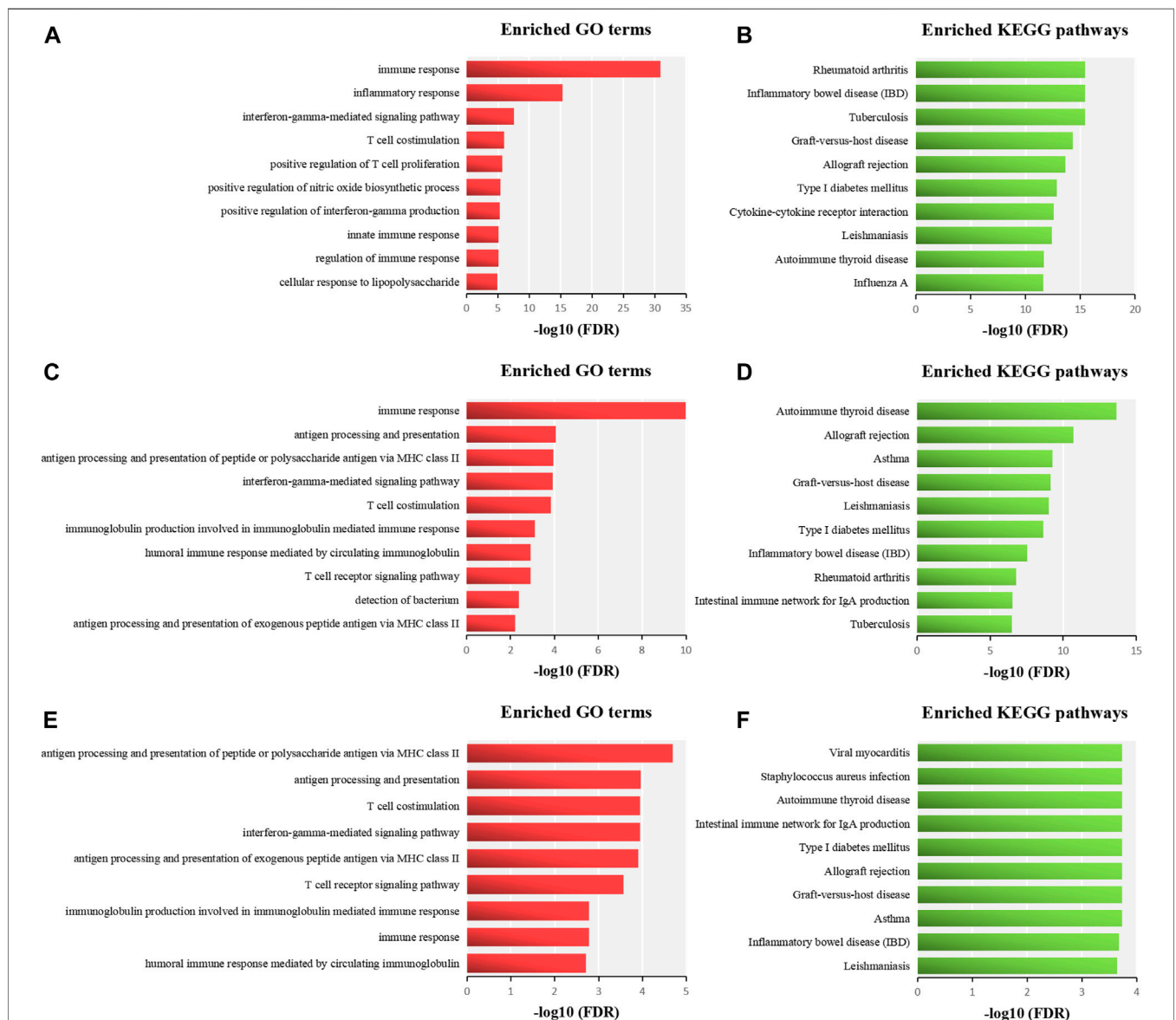


In this study, we calculated the similarity scores between 26 ADs by means of three similarity measurements. To ensure the accuracy of the subsequent analysis, we combined the results of NetSim, FunSim, and SemSim to evaluate all the AD pairs. We found ten potential pairs of similar ADs that were utilized to form a network containing an overall insight of the

information about AD-AD relationships and AD-gene relationships, which provided essential clues to understand the mechanisms shared by multiple ADs. Based on the AD pairs in this network, we detected three significant pairs of similar ADs (RA and SLE, MG and AIT, and AP and VKH), and then investigated the shared functional terms for each

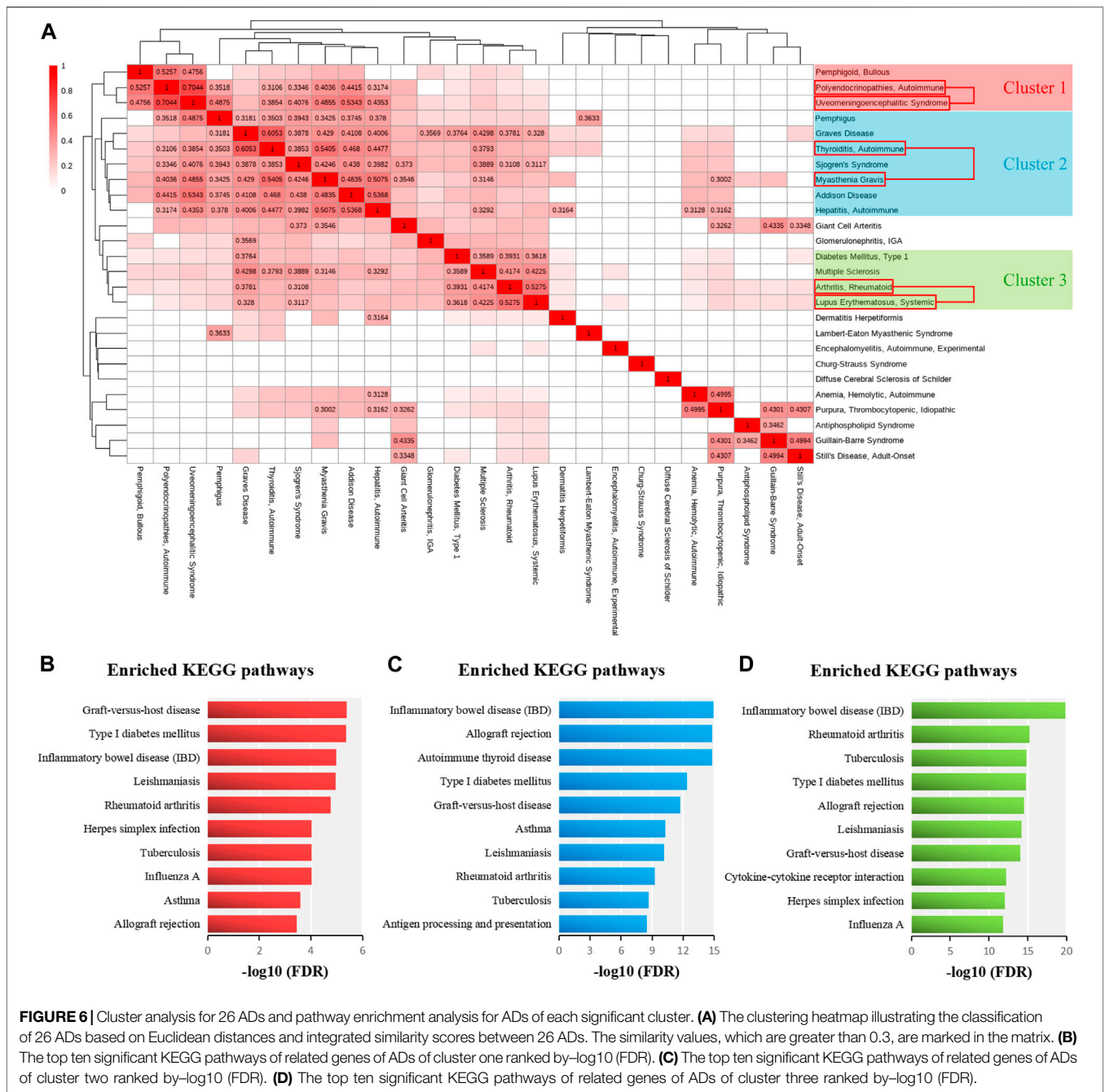
**TABLE 2 |** Top ten pairs of ADs ranked by integrated similarity scores.

Rank	Autoimmune disease	Autoimmune disease	Integrated similarity score
1	Polyendocrinopathies, Autoimmune	Uveomeningoencephalitic Syndrome	0.704362256
2	Thyroiditis, Autoimmune	Graves Disease	0.605323256
3	Myasthenia Gravis	Thyroiditis, Autoimmune	0.540495323
4	Addison Disease	Hepatitis, Autoimmune	0.53677382
5	Uveomeningoencephalitic Syndrome	Addison Disease	0.534304723
6	Arthritis, Rheumatoid	Lupus Erythematosus, Systemic	0.527496589
7	Pemphigoid, Bullous	Polyendocrinopathies, Autoimmune	0.52566119
8	Hepatitis, Autoimmune	Myasthenia Gravis	0.507462033
9	Anemia, Hemolytic, Autoimmune	Purpura, Thrombocytopenic, Idiopathic	0.499501
10	Guillain-Barre Syndrome	Still's Disease, Adult-Onset	0.499448795



**FIGURE 5 |** Functional enrichment analysis of related genes of RA and SLE (A-B), MG and AIT (C-D), and AP and VKH (E-F). Enriched functional terms are sorted in descending order according to their  $-\log_{10}(\text{FDR})$ , and the top ten significant GO terms in the BP and KEGG pathways of each significant pair of similar ADs are used for further analysis.





significant AD pair. We also employed cluster analysis on the integrated similarity matrix of 26 ADs to acquire some other ADs which were similar to significant AD pairs in genetics, and identified the risk genes which belonged to the same disease cluster. These results still need to be verified by more studies, but we hope that our observations can help researchers to dissect the complex pathogenesis of ADs.

By the functional enrichment analysis of 35 AD relevant genes, we mainly focused on GO terms involved in immune response, antigen processing and presentation, and  $IFN\gamma$ . The immune response is how the immune system defends against foreign

invaders, such as bacteria or viruses (Chaplin, 2010). ADs are triggered by aberrant immune response which damages healthy body part and is influenced by a large number of genes (Hill et al., 2008; Gregersen and Olsson, 2009). We concluded that the 35 genes might trigger a variety of ADs. On the other hand, the dysfunction of antigen processing and presentation might influence the emergence of ADs (Ritz and Seliger, 2001). In human bodies, antigens are processed into peptides of a certain length in association with major histocompatibility complex (MHC) molecules. T cells are capable of recognizing these fragmented peptides bound to the MHC to initiate immune responses (Purcell et al., 2016; Kelly and Trowsdale, 2019;

**TABLE 3 |** The shared causal genes among the ADs which belong to the same significant disease cluster.

Disease cluster	Shared gene
Cluster one	<i>HLA-DQA1</i>
Cluster two	<i>HLA-DQB1, HLA-DRB1</i>
Cluster three	<i>TNF, HLA-DRB1, PDCD1, PTPN22, CCR5, IL6, HLA-DQB1, CTLA4</i>

*Italics refers to gene symbols (gene names).*

Kotsias et al., 2019). As different ADs share the characteristic that risk is conferred by genes encoded within the MHC locus, antigen presentation generally seems to be crucial in ADs (Riedhammer and Weissert, 2015). For example, processing and presentation of self-antigens by different antigen presenting cells may result in MS (Stoeckle and Tolosa, 2010). Ultimately, IFN $\gamma$  is a pleiotropic cytokine secreted by immune cells and plays a critical role in innate and adaptive immunity (Tau and Rothman, 1999; Schoenborn and Wilson, 2007). Abnormal IFN $\gamma$  expression is correlated with considerable number of ADs. Although IFN $\gamma$  can mediate clearance of pathogenic insults, chronic exposure to IFN $\gamma$  is thought to cause many ADs, such as RA and SLE (Nielen et al., 2004; Lu et al., 2016). And the complex role of IFN $\gamma$  in ADs also has important therapeutic implications. Above evidences demonstrate that the three function aspects play important roles in AD-related mechanisms.

With regard to the three significant pairs of similar ADs, several studies have confirmed these similarity relationships. For example, (Wang et al., 2020) found that familial RA, SLE, and primary Sjögren's syndrome shared common genetic characteristics, and the genetic variations in T cell receptor signaling pathway genes which might become novel molecular targets for therapeutic interventions for the three ADs. (Liu et al., 2019) found that T cell receptor could become a promising diagnostic marker for RA and SLE. In addition, previous studies have confirmed that MG and AIT are similar in many aspects (Marino et al., 1997; Lopomo and Berrih-Aknin, 2017), and AIT frequently accompanies MG (Mao et al., 2011; Kubiszewska et al., 2016). The two diseases are both organ-specific ADs with a clear pathogenic effect of antibodies. Meanwhile, MG and AIT share the same predisposing genes (such as *PTPN22*, *CTLA4*, and *HLA*) and pathological mechanisms (such as T-cell immune-mediated mechanisms). Thus, we infer that AD genetic similarity research can help to explain the similar phenotypic and clinical features between ADs. These reports are consistent with our current results. Experimental studies on these AD pairs are desperately needed to provide important information to understand their intrinsic mechanisms. And further validation of these disease relationships in clinical trials will be a better option to turn them into clinical practice. Besides, the results of pathway enrichment analysis of related genes of significant pairs of similar ADs exposed possible PolyA. For example, the related genes of RA and SLE were enriched in pathways of IBD, T1D, and AITD. It was reported that AIT frequently coexisted with RA and SLE (Ordóñez-Canizares et al., 2020). Another study showed that AITD, RA, SLE, and IBD were observed in Sjögren's syndrome patients with PolyA (Amador-Patarroyo et al., 2012). And the related genes of MG and

AIT were enriched in pathways of T1D, IBD, and RA. Previous study found that the latent and overt PolyA in patients with AITD were associated with gastrointestinal, endocrinological, rheumatological, dermatological, and neurological ADs (Botello et al., 2020). The PolyA is not uncommon and multiple ADs that coexist in a single patient may share the same etiopathogenesis. Some genetic studies on ADs ignored the coexistence of other autoimmune conditions by implementing anachronistic nomenclature (i.e., primary or secondary ADs) (Rojas-Villarraga et al., 2012). We hope that researchers can take in account PolyA and concern whether or not patients have latent or overt PolyA in AD study.

With regard to the result of cluster analysis on 26 ADs, hitherto, a lot of reports have confirmed our viewpoint. For example, for the disease cluster two, a recent study found that chemokines were associated with the early phases of the autoimmune response in AIT, Graves disease, and Addison disease (Fallahi et al., 2020). For the disease cluster three, another study found major common gene expression changes at the target tissues of T1D, MS, RA, and SLE (Szymczak et al., 2021).

This study predicted AD pairs and clusters with high genetic similarity, as well as potential risk genes, biological processes, and pathways involved in multiple types of ADs. Despite the two diseases of a certain AD pair with high similarity score have different phenotypic or clinical features, they are likely to have similar or the same ways to elicit autoimmune responses in the human body. Consequently, we reason that similar ADs in genetics can be treated with similar therapeutics and drugs. We hope that these findings can aid in elucidating AD mechanisms, and provide more references for researchers.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

YD designed the study, analyzed the data, and drafted the manuscript. MC, JQ, and CW collected the data. QS, HR, and LL helped in analyzing the data, and calibrating the parameters of the algorithm. FZ and RZ provided ideas and revised the manuscript.

## FUNDING

This work was supported by the Natural Science Foundation of Heilongjiang Province (Grant No. LH 2019C043).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.758041/full#supplementary-material>

## REFERENCES

- Amador-Patarroyo, M. J., Arbelaez, J. G., Mantilla, R. D., Rodríguez-Rodríguez, A., Cárdenas-Roldán, J., Pineda-Tamayo, R., et al. (2012). Sjögren's Syndrome at the Crossroad of Polyautoimmunity. *J. Autoimmun.* 39, 199–205. doi:10.1016/j.jaut.2012.05.008
- Anaya, J.-M., Gómez, L., and Castiblanco, J. (2006). Is There a Common Genetic Basis for Autoimmune Diseases. *Clin. Develop. Immunol.* 13, 185–195. doi:10.1080/17402520600876762
- Anaya, J.-M. (2017). The Autoimmune Tautology. A Summary of Evidence. *Jt. Bone Spine* 84, 251–253. doi:10.1016/j.jbspin.2016.11.012
- Anaya, J.-M. (2014). The Diagnosis and Clinical Significance of Polyautoimmunity. *Autoimmun. Rev.* 13, 423–426. doi:10.1016/j.autrev.2014.01.049
- Barrat, F. J., Crow, M. K., and Ivashkiv, L. B. (2019). Interferon Target-Gene Expression and Epigenomic Signatures in Health and Disease. *Nat. Immunol.* 20, 1574–1583. doi:10.1038/s41590-019-0466-2
- Becker, K. G., Barnes, K. C., Bright, T. J., and Wang, S. A. (2004). The Genetic Association Database. *Nat. Genet.* 36, 431–432. doi:10.1038/ng0504-431
- Botello, A., Herrán, M., Salcedo, V., Rodríguez, Y., Anaya, J. M., and Rojas, M. (2020). Prevalence of Latent and Overt Polyautoimmunity in Autoimmune Thyroid Disease: A Systematic Review and Meta-analysis. *Clin. Endocrinol.* 93, 375–389. doi:10.1111/cen.14304
- Cárdenas-Roldán, J., Rojas-Villarraga, A., and Anaya, J.-M. (2013). How Do Autoimmune Diseases Cluster in Families? A Systematic Review and Meta-Analysis. *BMC Med.* 11, 73. doi:10.1186/1741-7015-11-73
- Chaplin, D. D. (2010). Overview of the Immune Response. *J. Allergy Clin. Immunol.* 125, S3–S23. doi:10.1016/j.jaci.2009.12.980
- Chavali, S., Barrenas, F., Kanduri, K., and Benson, M. (2010). Network Properties of Human Disease Genes with Pleiotropic Effects. *BMC Syst. Biol.* 4, 78. doi:10.1186/1752-0509-4-78
- Cheng, L., Li, J., Ju, P., Peng, J., and Wang, Y. (2014). SemFunSim: a New Method for Measuring Disease Similarity by Integrating Semantic and Gene Functional Association. *PLoS ONE* 9, e99415. doi:10.1371/journal.pone.0099415
- Cruz-Tapias, P., Pérez-Fernández, O. M., Rojas-Villarraga, A., Rodríguez-Rodríguez, A., Arango, M.-T., and Anaya, J.-M. (2012). Shared HLA Class II in Six Autoimmune Diseases in Latin America: A Meta-Analysis. *Autoimmune Dis.* 2012, 1–10. doi:10.1155/2012/569728
- Del Prete, E., Facchiano, A., and Liò, P. (2018). Bioinformatics Methodologies for Coeliac Disease and its Comorbidities. *Brief. Bioinform.* 21, 355–367. doi:10.1093/bib/bby109
- Dozmorov, M. G. (2019). Disease Classification: from Phenotypic Similarity to Integrative Genomics and beyond. *Brief. Bioinform.* 20, 1769–1780. doi:10.1093/bib/bby049
- Fallahi, P., Ferrari, S. M., Ragusa, F., Ruffilli, I., Elia, G., Paparo, S. R., et al. (2020). Th1 Chemokines in Autoimmune Endocrine Disorders. *J. Clin. Endocrinol. Metab.* 105, 1046–1060. doi:10.1210/clinem/dg2289
- Gregersen, P. K., and Olsson, L. M. (2009). Recent Advances in the Genetics of Autoimmune Disease. *Annu. Rev. Immunol.* 27, 363–391. doi:10.1146/annurev.immunol.021908.132653
- Hidalgo, C. A., Blumm, N., Barabási, A.-L., and Christakis, N. A. (2009). A Dynamic Network Approach for the Study of Human Phenotypes. *Plos Comput. Biol.* 5, e1000353. doi:10.1371/journal.pcbi.1000353
- Hill, N. J., King, C., and Flodstrom-Tullberg, M. (2008). Recent Acquisitions on the Genetic Basis of Autoimmune Disease. *Front. Biosci.* 13, 4838–4851. doi:10.2741/3043
- Hu, X., and Ivashkiv, L. B. (2009). Cross-regulation of Signaling Pathways by Interferon- $\gamma$ : Implications for Immune Responses and Autoimmune Diseases. *Immunity* 31, 539–550. doi:10.1016/j.immuni.2009.09.002
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res.* 37, 1–13. doi:10.1093/nar/gkn923
- Ji, J., Sundquist, J., and Sundquist, K. (2016). Gender-specific Incidence of Autoimmune Diseases from National Registers. *J. Autoimmun.* 69, 102–106. doi:10.1016/j.jaut.2016.03.003
- Kelly, A., and Trowsdale, J. (2019). Genetics of Antigen Processing and Presentation. *Immunogenetics* 71, 161–170. doi:10.1007/s00251-018-1082-2
- Kotsias, F., Cebrian, I., and Alloati, A. (2019). Antigen Processing and Presentation. *Int. Rev. Cel. Mol. Biol.* 348, 69–121. doi:10.1016/bs.ircmb.2019.07.005
- Kubiszewska, J., Szyluk, B., Szczudlik, P., Bartoszewicz, Z., Dutkiewicz, M., Bielecki, M., et al. (2016). Prevalence and Impact of Autoimmune Thyroid Disease on Myasthenia Gravis Course. *Brain Behav.* 6, e00537. doi:10.1002/brb3.537
- Lee, I., Blom, U. M., Wang, P. I., Shim, J. E., and Marcotte, E. M. (2011). Prioritizing Candidate Disease Genes by Network-Based Boosting of Genome-wide Association Data. *Genome Res.* 21, 1109–1121. doi:10.1101/gr.118992.110
- Li, P., Zheng, Y., and Chen, X. (2017). Drugs for Autoimmune Inflammatory Diseases: From Small Molecule Compounds to Anti-TNF Biologics. *Front. Pharmacol.* 8, 460. doi:10.3389/fphar.2017.00460
- Li, Y. R., Li, J., Zhao, S. D., Bradfield, J. P., Mentch, F. D., Maggadottir, S. M., et al. (2015). Meta-analysis of Shared Genetic Architecture across Ten Pediatric Autoimmune Diseases. *Nat. Med.* 21, 1018–1027. doi:10.1038/nm.3933
- Liu, X., Zhang, W., Zhao, M., Fu, L., Liu, L., Wu, J., et al. (2019). T Cell Receptor  $\beta$  Repertoires as Novel Diagnostic Markers for Systemic Lupus Erythematosus and Rheumatoid Arthritis. *Ann. Rheum. Dis.* 78, 1070–1078. doi:10.1136/annrheumdis-2019-215442
- Lopomo, A., and Berrih-Aknin, S. (2017). Autoimmune Thyroiditis and Myasthenia Gravis. *Front. Endocrinol.* 8, 169. doi:10.3389/fendo.2017.00169
- Lu, R., Munroe, M. E., Guthridge, J. M., Bean, K. M., Fife, D. A., Chen, H., et al. (2016). Dysregulation of Innate and Adaptive Serum Mediators Precedes Systemic Lupus Erythematosus Classification and Improves Prognostic Accuracy of Autoantibodies. *J. Autoimmun.* 74, 182–193. doi:10.1016/j.jaut.2016.06.001
- Mao, Z.-F., Yang, L.-X., Mo, X.-A., Qin, C., Lai, Y.-R., He, N.-Y., et al. (2011). Frequency of Autoimmune Diseases in Myasthenia Gravis: a Systematic Review. *Int. J. Neurosci.* 121, 121–129. doi:10.3109/00207454.2010.539307
- Margo, C. E., and Harman, L. E. (2016). Autoimmune Disease: Conceptual History and Contributions of Ocular Immunology. *Surv. Ophthalmol.* 61, 680–688. doi:10.1016/j.survophthal.2016.04.006
- Marinó, M., Ricciardi, R., Pinchera, A., Barbesino, G., Manetti, L., Chiovato, L., et al. (1997). Mild Clinical Expression of Myasthenia Gravis Associated with Autoimmune Thyroid Diseases. *J. Clin. Endocrinol. Metab.* 82, 438–443. doi:10.1210/jcem.82.2.3749
- McGinness, J. L., Bivens, M.-M. C., Greer, K. E., Patterson, J. W., and Saulsbury, F. T. (2006). Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked Syndrome (IPEX) Associated with Pemphigoid Nodularis: a Case Report and Review of the Literature. *J. Am. Acad. Dermatol.* 55, 143–148. doi:10.1016/j.jaad.2005.08.047
- Nielen, M. M. J., van Schaardenburg, D., Reesink, H. W., van de Stadt, R. J., van der Horst-Bruinsma, I. E., de Koning, M. H. M. T., et al. (2004). Specific Autoantibodies Precede the Symptoms of Rheumatoid Arthritis: a Study of Serial Measurements in Blood Donors. *Arthritis Rheum.* 50, 380–386. doi:10.1002/art.20018
- Ordoñez-Cañizares, M. C., Mena-Vázquez, N., Redondo-Rodríguez, R., Manrique-Arija, S., Jimenez-Núñez, F. G., Ureña-Garnica, I., et al. (2020). Frequency of Polyautoimmunity in Patients with Rheumatoid Arthritis and Systemic Lupus Erythematosus. *J. Clin. Rheumatol.* Publish Ahead of Print. doi:10.1097/RHU.0000000000001574
- Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., Jonnalagadda, C. K., Surendranath, V., et al. (2003). Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans. *Genome Res.* 13, 2363–2371. doi:10.1101/gr.1680803
- Place, D. E., and Kanneganti, T.-D. (2020). The Innate Immune System and Cell Death in Autoinflammatory and Autoimmune Disease. *Curr. Opin. Immunol.* 67, 95–105. doi:10.1016/j.coi.2020.10.013
- Purcell, A. W., Croft, N. P., and Tschärke, D. C. (2016). Immunology by Numbers: Quantitation of Antigen Presentation Completes the Quantitative Milieu of Systems Immunology!. *Curr. Opin. Immunol.* 40, 88–95. doi:10.1016/j.coi.2016.03.007
- Quintero-Ronderos, P., and Montoya-Ortiz, G. (2012). Epigenetics and Autoimmune Diseases. *Autoimmune Dis.* 2012, 1–16. doi:10.1155/2012/593720
- Richard-Miceli, C., and Criswell, L. A. (2012). Emerging Patterns of Genetic Overlap across Autoimmune Disorders. *Genome Med.* 4, 6. doi:10.1186/gm305
- Riedhammer, C., and Weissert, R. (2015). Antigen Presentation, Autoantigens, and Immune Regulation in Multiple Sclerosis and Other Autoimmune Diseases. *Front. Immunol.* 6, 322. doi:10.3389/fimmu.2015.00322

- Ritz, U., and Seliger, B. (2001). The Transporter Associated with Antigen Processing (TAP): Structural Integrity, Expression, Function, and its Clinical Relevance. *Mol. Med.* 7, 149–158. doi:10.1007/bf03401948
- Rojas-Villarraga, A., Amaya-Amaya, J., Rodriguez-Rodriguez, A., Mantilla, R. D., and Anaya, J.-M. (2012). Introducing Polyautoimmunity: Secondary Autoimmune Diseases No Longer Exist. *Autoimmune Dis.* 2012, 1–9. doi:10.1155/2012/254319
- Rosenblum, M. D., Gratz, I. K., Paw, J. S., and Abbas, A. K. (2012). Treating Human Autoimmunity: Current Practice and Future Prospects. *Sci. Transl. Med.* 4, 125sr121. doi:10.1126/scitranslmed.3003504
- Schoenborn, J. R., and Wilson, C. B. (2007). Regulation of Interferon- $\gamma$  during Innate and Adaptive Immune Responses. *Adv. Immunol.* 96, 41–101. doi:10.1016/S0065-2776(07)96002-2
- Schriml, L. M., Mitraka, E., Munro, J., Tauber, B., Schor, M., Nickle, L., et al. (2019). Human Disease Ontology 2018 Update: Classification, Content and Workflow Expansion. *Nucleic Acids Res.* 47, D955–D962. doi:10.1093/nar/gky1032
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303
- Stoeckle, C., and Tolosa, E. (2009). Antigen Processing and Presentation in Multiple Sclerosis. *Results Probl. Cel Differ.* 51, 149–172. doi:10.1007/400\_2009\_22
- Stojanovich, L., and Marisavljevic, D. (2008). Stress as a Trigger of Autoimmune Disease. *Autoimmun. Rev.* 7, 209–213. doi:10.1016/j.autrev.2007.11.007
- Szymczak, F., Colli, M. L., Mamula, M. J., Evans-Molina, C., and Eizirik, D. L. (2021). Gene Expression Signatures of Target Tissues in Type 1 Diabetes, Lupus Erythematosus, Multiple Sclerosis, and Rheumatoid Arthritis. *Sci. Adv.* 7. doi:10.1126/sciadv.abd7600
- Tau, G., and Rothman, P. (1999). Biologic Functions of the IFN- $\gamma$  Receptors. *Allergy* 54, 1233–1251. doi:10.1034/j.1398-9995.1999.00099.x
- Ueda, H., Howson, J. M. M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., et al. (2003). Association of the T-Cell Regulatory Gene CTLA4 with Susceptibility to Autoimmune Disease. *Nature* 423, 506–511. doi:10.1038/nature01621
- Wang, Y., Chen, S., Chen, J., Xie, X., Gao, S., Zhang, C., et al. (2020). Germline Genetic Patterns Underlying Familial Rheumatoid Arthritis, Systemic Lupus Erythematosus and Primary Sjögren's Syndrome Highlight T Cell-Initiated Autoimmunity. *Ann. Rheum. Dis.* 79, 268–275. doi:10.1136/annrheumdis-2019-215533
- Wieber, K., Zimmer, C. L., and Hertl, M. (2021). Detection of Autoreactive CD4+ T Cells by MHC Class II Multimers in HLA-Linked Human Autoimmune Diseases. *J. Clin. Invest.* 131 (9), e148674. doi:10.1172/JCI148674
- Zhang, S.-B., and Lai, J.-H. (2016). Exploring Information from the Topology beneath the Gene Ontology Terms to Improve Semantic Similarity Measures. *Gene* 586, 148–157. doi:10.1016/j.gene.2016.04.024
- Zhang, S.-B., and Lai, J.-H. (2015). Semantic Similarity Measurement between Gene Ontology Terms Based on Exclusively Inherited Shared Information. *Gene* 558, 108–117. doi:10.1016/j.gene.2014.12.062

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ding, Cui, Qian, Wang, Shen, Ren, Li, Zhang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.