DATABASE

GAAD: A Gene and Autoimmune Disease Association Database

Guanting Lu 1,#,a, Xiaowen Hao 2,#,b, Wei-Hua Chen 2,#,c, Shijie Mu 1,#,d

1 Department of Blood Transfusion, Tangdu Hospital, Fourth Military Medical University, Xi’an 710032, China
2 MOE Key Laboratory of Molecular Biophysics, Hubei Key Laboratory of Bioinformatics and Molecular-imaging, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

Received 3 December 2017; revised 4 April 2018; accepted 22 May 2018
Available online 27 September 2018

Handled by Quan-Zhen Li

KEYWORDS

Autoimmune diseases; Disease–gene association; Database; Text mining

Abstract  Autoimmune diseases (ADs) arise from an abnormal immune response of the body against substances and tissues normally present in the body. More than a hundred of ADs have been described in the literature so far. Although their etiology remains largely unclear, various types of ADs tend to share more associated genes with other types of ADs than with non-AD types. Here we present GAAD, a gene and AD association database. In GAAD, we collected 44,762 associations between 49 ADs and 4249 genes from public databases and MEDLINE documents. We manually verified the associations to ensure the quality and credibility. We reconstructed and recapitulated the relationships among ADs using their shared genes, which further validated the quality of our data. We also provided a list of significantly co-occurring gene pairs among ADs; with embedded tools, users can query gene co-occurrences and construct customized co-occurrence network with genes of interest. To make GAAD more straightforward to experimental biologists and medical scientists, we extracted additional information describing the associations through text mining, including the putative diagnostic value of the associations, type and position of gene polymorphisms, expression changes of implicated genes, as well as the phenotypical consequences, and grouped the associations accordingly. GAAD is freely available at http://gaad.medgenius.info.

* Corresponding authors.
E-mail: weihuachen@hust.edu.cn (Chen WH), musj1963@fmmu.edu.cn (Mu S).
# Equal contribution.
a ORCID: 0000-0002-8556-8687.
b ORCID: 0000-0002-0175-1186.
c ORCID: 0000-0001-5160-4398.
d ORCID: 0000-0003-1267-6000.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

https://doi.org/10.1016/j.gpb.2018.05.001
1672-0229 © 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
An autoimmune disease (AD) is a condition that arises from an abnormal immune response to a normal body part. ADs may be restricted to certain organs or involve a particular tissue in different places. For example, autoimmune thyroiditis only affects the thyroid glands, while Goodpasture's disease can affect the basement membrane in both the lung and the kidney. Many ADs, with the exception of rheumatoid arthritis (RA) and autoimmune thyroiditis, are individually rare; but together they affect more than 10% of the population worldwide, with annual net increase of ~19% in incidents. It has been estimated in the United States alone that up to 23.5 million Americans suffer from in total 24 ADs, accounting for ~7.5% of the whole population (see also https://www.aard.org, the American Autoimmune Related Diseases Association for the latest statistics), and the annual health care costs for ADs are in the range of $100 billion (data from https://www.womenshealth.gov/a-z-topics/autoimmune-diseases; accessed on July 19, 2017). In comparison, cancer affects up to 9 million Americans and costs ~$57 billion (data source: American Cancer Association; accessed on July 19, 2017).

Similar to other complex diseases such as cancer, many genes have been reported to be associated with ADs, but together they only account for a minor proportion of the disease risks. For example, a recent fine-mapping study using 33,595 patients with inflammatory bowel disease (IBD) estimates that ~22%–28% of the risks can be explained by in total 94 genomic regions. Previous observations in monogenetic twins indicated that genetic influences only account for ~25%–40% of the disease risks. These results indicate that either there are hidden factors such as epigenetics and/or protein post-translational modifications (PTMs) that have not been included in their experimental design, or environmental influences and/or gene–environment interactions constitute the predominant factor for disease risks. For example, it has been shown that the gut microbiota–host interactions have been reported to significantly shape the genetic architecture of some ADs, such as IBD, lupus nephritis, RA, multiple sclerosis, and Graves’ disease. Nonetheless, due to the diverse and difficult-to-quantify nature of environmental factors, genetic background and DNA mutations are still the major measurable disease risks for ADs so far.

Although the triggering factors for many ADs remain unknown, genome-wide association studies (GWAS) have identified many nucleotide polymorphisms, with some of them likely to be causal, and associated risk loci. Not surprisingly, many ADs share associated risk loci with other types of ADs significantly more than with other disease types, due to their shared triggering factors, and many ADs could co-occur in the same patients. In addition, risk loci in ADs often act independently, with an additive effect on the disease outcomes.

Disease-specific databases are useful for researchers to find and retrieve the appropriate information. In recent years, the number of biomedical databases has increased dramatically. However, a gene–disease association database dedicated to ADs is still unavailable. In this study, we presented GAAD, a gene–disease association database for ADs, aimed to enhance our understanding of ADs. The following information is collected into a central database: (i) a comprehensive gene catalog associated with ADs, facilitating genome-wide analyses as well as gene-centered biological studies; (ii) relevant features describing the characteristics of such associations, including the putative diagnostic values of the association, time of disease onset due to dysfunction of an associated gene and expression changes of causal and/or implicated genes during disease progression, allowing quick selection of suitable biomarkers by experimental and/or clinical researchers. In addition, we also obtain a disease network using co-occurring gene pairs commutated from our collected data, allowing researchers to compare disease similarity and get hints on novel causal genes of ADs. For example, if most of the known causal genes of an AD are associated with another AD, then other genes associated with the latter disease could also be associated with the former disease.

In GAAD, we collected in total 44,763 associations between 4249 genes and 49 ADs, each association as an independent piece of evidence supporting the link between a gene and a disease. We obtained the association from public databases including the NCBI Gene database and GeneCards, literature describing large-scale GWAS analyses, and text mining results from MEDLINE documents. We manually verified the associations to ensure the data quality. Additionally, we extracted available information describing the associations through text mining, including the putative diagnostic value of the associations, type and position of gene polymorphisms, expression changes of implicated genes, and the phenotypical consequences, and grouped the associations accordingly, allowing experimental biologists and medical scientists to quickly and easily find relevant information to genes of interest.

GAAD is implemented with Javascript, HTML, PHP, and AngularJS (a model-view-controller framework for creating Javascript web applications; https://angularjs.org). GAAD uses Bootstrap for most of its front-end (i.e., the user interface) components and adopts MySQL as a relational database management system to store all the data. All codes are developed using PhpStorm and educational version. The whole website is hosted on a CentOS Educational version (one of the most popular Linux distributions; https://www.centos.org) operating system.

**Database content and usage**

**Disease selection**

We took the disease list from the American Autoimmune Related Diseases Association, cross-referenced with the list of diseases included in NCBI. Consequently, 49 ADs, for which gene associations can be found by our data collection process (see below), were retained.

**Data acquisition**

We performed text mining analysis using a customized Python pipeline by searching for co-occurrences of genes (e.g., gene...
symbol, locus ID, common name, alias, and variant) and diseases of interest from the titles or abstracts of MEDLINE documents available from the NCBI PubMed database (https://www.ncbi.nlm.nih.gov/pubmed). The criteria include: (1) the gene name and the disease name should be in the same sentence; (2) this sentence should also contain one of the following key words/phrases or their aliases: aberrant, account for, altered, associated, caused, confer, contribute, curb, downregulated, elevate, evoked, higher, implicated, increase, induce, influence, interact, involved, lead to, link, mediate, modulate, overexpressed, prevent, protective, reduce, regulate, related, relationship, treat, biomarker, clinic trial, therapy, diagnose, risk factor, target, prognostic, pathogenic, pathogenesis, progression, and predisposing factor. In addition, we also extracted features describing the relationships between genes and ADs (see above) from the identified sequences. We obtained 135,673 associations from 90,918 MEDLINE documents through initial text mining. The retrieved information was further verified by three rounds of manual inspection.

We also searched for gene–disease associations in the NCBI Gene database using disease names (and aliases) as input. The retrieved entries were also manually curated and incorporated into GAAD.

In addition, we used an in-house Perl script to search and retrieve gene–disease associations from the GeneCards database (http://www.genecards.org), which uses a disease relevance score to describe the credit of each association. The relevance score of an association was calculated by comparing “the observed number of MEDLINE documents where both elements appear together and the number of MEDLINE documents where both appear independently” with an expected value based on a hypergeometric distribution, which reflects an order of importance, with higher disease relevance score indicating more co-occurrences.

Data analyses

We calculated gene co-occurrences for all possible pairs using Fisher’s exact test to count data (the fisher.test() function)
implemented in R (https://cran.r-project.org). This R function takes a 2\times 2 contingency table as input; the data include four values which represent the numbers of ADs associated with (1) both genes in the pair, (2) only one gene, (3) only the other gene and (4) none of them. We reported significantly co-occurring gene pairs in GAAD.

For the two genes in a co-occurring pair, we also tested whether they are in the same KEGG pathways using R modules including pathview [18], KEGGREST (an R package utilizing KEGG Representational State Transfer (REST) programmable interface to access KEGG data; available at https://bioconductor.org/packages/release/bioc/html/KEGGREST.html), and gage (https://bioconductor.org/packages/release/bioc/html/gage.html). Genes in the same pathways are likely to have similar and/or relevant functions. However, it does not mean that two genes are not functionally similar if they are not in the same pathway, since only small fractions of human genes (6356 in this case) are involved in various pathways.

To test whether co-occurring genes are more likely to share similar functions, we also performed an additional analysis using Gene Ontology (GO), since GO covers more human proteins/genes than the KEGG database. We used a R package GOSemSim [19] to perform such analysis. GOSemSim computes semantic similarity among GO terms and reports a similarity score for each input gene pair ranging from 0 to 1, where 0 indicates no similarity and 1 indicates at least one GO term is shared by two genes (although it does not mean their functions are identical).

### Database contents and organization

The main contents of our database are organized into two pages, namely the “DISEASES” page and the “GENES” page.

| Autoimmune disease          | No. of associated genes |
|------------------------------|-------------------------|
| Rheumatoid arthritis         | 1324                    |
| Multiple sclerosis           | 977                     |
| Systemic lupus erythematosus | 876                     |
| Inflammatory bowel disease   | 843                     |
| Crohn’s disease              | 788                     |
| Psoriasis                    | 709                     |
| Ulcerative colitis           | 680                     |
| Parkinson’s disease          | 475                     |
| Systemic sclerosis           | 469                     |
| Atopic dermatitis            | 394                     |

*Note: A complete and up-to-date list of autoimmune diseases and the number of associated genes is available at http://gaad.medgenius.info/diseases/*.

Figure 2 The number of associated diseases for the 4249 genes follows a power-law distribution

Overall the distribution of the number of diseases that a gene is associated with follows the power-law distribution. A large number of genes are associated with a few diseases, while a few genes are associated with a large number of diseases. Values on both axes are log-transformed.
At the beginning of “DISEASES” (“GENES”) page, a summary table lists available diseases (genes) and the number of associated genes (diseases). We also provide widgets that allow users to search for or filter out diseases (genes) by user-supplied keywords. After clicking on a disease (gene) name, users will be redirected to a new page containing detailed information for the disease (gene), including a list of associated genes (diseases) and detailed evidence supporting the associations.

Most of our supporting evidence was obtained through text mining, we thus include the sentences in which the disease–gene association was described in the evidence view. We also highlight different types of keywords with different font and background colors (Figure 1). Text mining results are often error-prone; we thus carried out multi rounds of aggressive manual curation to remove any possible false results. However, despite these efforts, there might still be small fraction of false positive results. We thus include a user-feedback mechanism, allowing users to vote on each text mining entry. As shown in Figure 1, users can use the thumb-up and thumbs-down buttons to vote on the creditability of the entry; the background color of the table cell will be changed automatically according to the voting results.

We collected in total 41,235 associations from 19,299 MEDLINE documents through text mining followed by multi rounds of manual curation (see Data acquisition). We obtained additional 3489 associations from public databases including NCBI Gene and GeneCards. Together, we collected in total 44,762 associations between the 49 ADs and 4249 genes in our database.

The number of genes associated with an AD ranges from 1324 (RA) to 2 (vitiligo-associated multiple autoimmune disease), as shown in Table 1. Out of the associated genes in GAAD, more than half of the genes are associated with only one disease. Overall, the distribution of the number of diseases which a gene is associated with follows the power-law distribution (data from https://en.wikipedia.org/wiki/power_law; accessed on July 20, 2017) (Figure 2). In total 33 genes are associated with more than 20 ADs (Table 2); not surprisingly, most genes are involved in immune response.

It should be pointed out that in this study we adopted a broad definition of “gene”, which refers to not only any discrete locus of heritable, genomic sequence that affect an organism’s traits, but also its products including RNAs and proteins.

**Genes shared by ADs and significantly co-occurring gene pairs**

ADs collected in our database share significant proportion of their associated genes with other types of ADs. For example, for any given AD, it shares at least one associated gene with another AD. Therefore a network based on shared genes

---

**Table 2 Genes associated with > 20 autoimmune diseases in GAAD**

| Gene ID in GAAD | No. of associated diseases | Gene symbol | Protein encoded |
|----------------|---------------------------|-------------|----------------|
| 7124           | 34                        | TNF         | Tumor necrosis factor |
| 3569           | 29                        | IL6         | Interleukin 6 |
| 920            | 29                        | CD4         | CD4 molecule |
| 3586           | 28                        | IL10        | Interleukin 10 |
| 3558           | 28                        | IL2         | Interleukin 2 |
| 3123           | 28                        | HLA-DRB1    | Major histocompatibility complex, class II, DR beta 1 |
| 3605           | 27                        | IL17A       | Interleukin 17A |
| 3548           | 27                        | IFNG        | Interferon, gamma |
| 59043          | 26                        | FOXP3       | Forkhead box P3 |
| 3565           | 26                        | IL4         | Interleukin 4 |
| 3627           | 25                        | CXCL10      | Chemokine (C-X-C motif) ligand 10 |
| 3553           | 25                        | IL18        | Interleukin 1, beta |
| 51497          | 24                        | NELFCD      | Negative elongation factor complex member C/D |
| 1493           | 24                        | CTLA4       | Cytotoxic T-lymphocyte-associated protein 4 |
| 925            | 24                        | CD8A        | CD8a molecule |
| 7422           | 23                        | VEGFA       | Vascular endothelial growth factor A |
| 7099           | 23                        | TLR4        | Toll-like receptor 4 |
| 7040           | 23                        | TGFB1       | Transforming growth factor, beta 1 |
| 3606           | 23                        | IL18        | Interleukin 18 |
| 51561          | 22                        | IL23A       | Interleukin 23, alpha subunit p19 |
| 3576           | 22                        | CXCL8       | Chemokine (C-X-C motif) ligand 8 |
| 1401           | 22                        | CRP         | C-reactive protein, pentraxin-related |
| 6347           | 21                        | CCL2        | Chemokine (C-C motif) ligand 2 |
| 4524           | 21                        | MTHFR       | Methyltetrahydrofolate reductase (NADPH) |
| 4282           | 21                        | MIF         | Macrophage migration inhibitory factor (glycosylation-inhibiting factor) |
| 3630           | 21                        | INS         | Insulin |
| 3119           | 21                        | HLA-DQ61    | Major histocompatibility complex, class II, DQ beta 1 |
| 3106           | 21                        | HLA-B       | Major histocompatibility complex, class I, B |
| 720            | 21                        | C4A         | Complement component 4A (Rodgers blood group) |
| 90865          | 20                        | IL33        | Interleukin 33 |
| 50616          | 20                        | IL22        | Interleukin 22 |
| 3559           | 20                        | IL2RA       | Interleukin 2 receptor, alpha |
| 3383           | 20                        | ICAM1       | Intercellular adhesion molecule 1 |
| 959            | 20                        | CD40LG      | CD40 ligand |

Note: A complete and up-to-date list of genes associated with > 20 autoimmune diseases is available at http://gaad.medgenius.info/genes/.
connects all our collected ADs (Figure 3, left panel); on average, an AD shares ~30% of its associated genes with other types of ADs.

To allow users to explore the relationships among ADs and their shared genes, at the bottom of the “DISEASES” page, we provide an interactive view for the relationships. As shown in Figure 3 (left panel), circles (nodes) represent ADs, with their diameters being proportional to the numbers of associated genes; the connecting lines (edges) represent their relationships, with the line width being proportional to the numbers of shared genes. Mouseover of a node highlights all its connecting edges and direct neighbors in the network (Figure 3, left panel). In addition, a list of its connected ADs is also shown (Figure 3, right panel). The nodes and edges are clickable; when clicked, the users will be redirected to new pages with detailed information on the AD and the shared genes between two ADs, respectively.

In addition, we also provide a list of significantly co-occurring gene pairs in GAAD. In total we have identified 5440 significantly co-occurring gene pairs, which can be found on the “GENES” page. At the bottom of the page, there is also an interactive visualization widget, allowing users to explore the gene co-occurrence network constructed using these gene pairs.

We queried the KEGG database to see whether the two genes of a pair tend to be involved in the same metabolic or signaling pathway, and/ or share functional similarities according to their annotations in the GO database. Compared to the randomly generated and non-significant gene pairs, significantly

![Figure 3](Image)
co-occurring pairs tend to have similar functions ($P < 2E^{-16}$ for both KEGG and GO analyses, Fisher’s exact test). In addition, we found that the more ADs in which both genes co-occur, the higher chance that the two genes are in the same pathway (Figure 5A) and share similar functions (Figure 5B).

**Extraction of clinically-relevant information for associations**

To make GAAD more straightforward to experimental biologists and medical scientists, we extracted clinically-relevant features describing the gene–disease associations from the text mining results, and assigned them into “tag groups”. In total ten tag groups were created, each describing different aspects of the associations. These tag groups include *keyword of diagnostic value, onset stage, population* (in which the study was performed), *animal model, expression change* (between control and patient groups), *post-translational modification, associated microRNA, keyword* (used to do the query), *mutation type*, and *mutation position*. The “tag groups” are available at http://gaad.medgenius.info/tags/.

**Example usage 1: Searching for putative causal genes for ADs**

‘Tags’ are essentially keywords characterizing many aspects of the associations between genes and ADs, and are only available for text mining results. We grouped ‘tags’ into ten tag groups according to the characteristics of the corresponding associations, for example, population in which the association was described, the mutation type and position of the associated gene, and whether any potential diagnostic value of the association has been described. Information about the diagnostic value of an association would be most interesting
**Figure 5** Genes in a co-occurring pair often have similar functions

The more ADs in which both genes of a pair co-occurring, the higher chance that they are in the same pathway (A) and functionally similar (B); see “Data analysis” for details on the calculation of functional similarities. Random refers to randomly generated pairs. NS pairs1 indicates gene pairs that do not co-occur in any ADs, whereas NS pairs2 indicates gene pairs co-occur in ADs but the co-occurrences are not significant ($P \geq 0.5$, Fisher’s exact test; see “Data analysis” for details). The dots at the bottom of bins (4, 5] and (6, 10] are outliers.

**Figure 6** Searching for putative causal genes for ADs

Following the steps indicated on the screenshot, users are able to find putative causal genes for ADs collected in our database. The resulting entries (mostly are PubMed records) are organized according to the implicated diseases. Here “Tag category” refers to the major categories (i.e., manually curated biologically meaningful categories) such as “Keyword of diagnostic value”, “Mutation type” and “Onset stage”, while “Sub category” refers to keywords belonging to the major category.
to clinicians. To find such information, users can simply go to the “TAGS” page, select the “Keywords of diagnostic value” from the “Tag category” dropdown menu. A second dropdown menu containing sub-categories is also available for users to further limit the number of matched entries (as shown in Figure 6). By default, the first sub-category from the second dropdown menu is always selected to avoid possible server overload due to the large number of matching entries under the corresponding “Tag category”.

Example usage 2: analyzing gene co-occurrences in ADs with user-supplied genes

Gene pairs co-occurring in multiple ADs often share functional similarities and tend to be involved in the same metabolic or signaling pathways. On the “GENES” page we provide a list of pre-commutated co-occurring pairs and a network constructed with gene pairs co-occurring in more than 20 ADs. Using this module, users can also construct their own

![Gene co-occurrence network](image)

Figure 7  Analyzing gene co-occurrences in ADs with user-supplied genes
At the bottom of the “GENES” page, users can enter a list of genes of interest into the “User input genes” textbox to generate a gene co-occurrence network for these genes. Shown here is a network constructed from the top 10 genes that are associated with more than 25 ADs collected in our database. The 10 genes are TNF, CD4, IL6, HLA-DRB1, IL2, IL10, IFNG, IL17A, IL4, and FOXP3. Selected node (“CD4” in this case) will be outlined in yellow with its directly connected edges highlighted in yellow as well. Line widths are proportional to the numbers of co-occurring diseases between genes and edges are also color coded to show the same information, with darker gray indicating fewer co-occurring diseases and darker red indicating more co-occurring diseases.
gene co-occurrence network with a list of user-supplied genes. As shown in Figure 7, we submitted the ten genes that are associated with more than 25 ADs in GAAD and constructed a co-occurrence network. These genes are often interconnected within this network, suggesting that many of them are also likely to be associated with the same ADs.

Perspectives and concluding remarks

In this study we described GAAD, a disease–gene association database for autoimmune diseases. We assembled data from public databases and MEDLINE documents. We verified the associations by several rounds of manual inspection, especially for those derived from text mining to ensure the quality and credibility. To help users to better understand the interrelationships among ADs, we calculated shared genes among ADs and significantly co-occuring gene pairs. We provide powerful and easy-to-use web interface for users to access these data; we also include embedded tools so that users can query gene co-occurrences and construct customized co-occurrence network with genes of interest. In addition, associations were grouped according to relevant information, including the putative diagnostic value of the associations, type and position of gene polymorphisms, expression changes of implicated genes and the phenotypical consequences; such information should be of interest to experimental biologists and medical scientists. GAAD is freely available at http://gaad.medgenius.info.

Authors’ contributions

GL, WC, and SM conceived the study, and calculated gene co-occurrences for all possible pairs; WC and XH designed the website. GL, XH, WC, and SM performed text mining and searched for gene–disease associations. XH drafted the manuscript; GL and WC edited the final manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

We would like to thank S. Balakrishnan, a member of our laboratory, for proofreading the manuscript.

References

[1] Davidson A, Diamond B. Autoimmune diseases. N Engl J Med 2001;345:340–50.
[2] Dayan CM, Daniels GH. Chronic autoimmune thyroiditis. N Engl J Med 1996;335:99–107.
[3] Salama AD, Levy JB, Lightstone L, Pusey CD. Goodpasture’s disease. Lancet 2001;358:917–20.
[4] Lerner A, Jeremias P, Matthias T. The world incidence and prevalence of autoimmune diseases is increasing. Int J Celiac Dis 2015;5:151–5.
[5] Huang H, Fang M, Jostins L, Umicevic Mirkov M, Boucher G, Anderson CA, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature 2017;547:173–8.
[6] Karopka T, Fluck J, Mevissen HT, Glass A. The Autoimmune Disease Database: a dynamically compiled literature-derived database. BMC Bioinformatics 2006;7:325.
[7] Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119–24.
[8] Mu Q, Zhang H, Liao X, Lin K, Liu H, Edwards MR, et al. Control of lupus nephritis by changes of gut microbiota. Microbiome 2017;5:73.
[9] Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med 2015;21:895–905.
[10] Maeda Y, Takeda K. Role of gut microbiota in rheumatoid arthritis. J Clin Med 2017;6:60.
[11] Shahi SK, Freedman SN, Mangalam AK. Gut microbiome in multiple sclerosis: the players involved and the roles they play. Gut Microbes 2017;8:607–15.
[12] Kohling HL, Plummer SF, Marchesi JR, Davidge KS, Ludgate M. The microbiota and autoimmunity: their role in thyroid autoimmune diseases. Clin Immunol 2017;183:63–74.
[13] Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM, et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat Genet 2016;48:1418–24.
[14] Boelaert K, Newby PR, Simmonds MJ, Holder RL, Carr-Smith JD, Heward JM, et al. Prevalence and relative risk of other autoimmune diseases in subjects with autoimmune thyroid disease. Am J Med 2010;123:183.e1–9.
[15] Henderson RD, Bain CJ, Pender MP. The occurrence of autoimmune diseases in patients with multiple sclerosis and their families. J Clin Neurosci 2000;7:434–7.
[16] Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 2016;54:1.30.1–33.
[17] MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 2017;45: D896–901.
[18] Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-data integration and visualization. Bioinformatics 2013;29:1830–1.
[19] Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 2010;26:976–8.