RADB: a database of rheumatoid arthritis-related polymorphisms

Ruijie Zhang,1,*†, Meiwei Luan1,†, Zhenwei Shang1,†, Lian Duan1,†, Guoping Tang2,†, Miao Shi1, Wenhua Lv1, Hongjie Zhu1, Jin Li1, Hongchao Lv1, Mingming Zhang1, Guiyou Liu3, He Chen4,* and Yongshuai Jiang1,*

1College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China, 2Yiwu Hospital, Zhejiang University, Yiwu 322000, China, 3Genome Analysis Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China, 4Department of Pathology, Harbin Medical University, Harbin 150086, China

*Corresponding author: Tel: +86045186620941-126; Fax: +86045186615922; Email: jiangyongshuai@gmail.com
*Correspondence may also be addressed to Ruijie Zhang. Tel: +86045186650721-106; Fax: +86045186615922; Email: zhangruijie2013@gmail.com and He Chen. Tel: +86045186650721-106; Fax: +86045186615922; Email: chenhe201406@163.com
†These authors contributed equally to this work.


Received 16 December 2013; Revised 13 August 2014; Accepted 19 August 2014

Abstract

Rheumatoid arthritis (RA) is an autoimmune disease that has a complex genetic basis. Therefore, it is important to explore the genetic background of RA. The extensive recent application of polymorphic genetic markers, especially single nucleotide polymorphisms, has presented us with a large quantity of genetic data. In this study, we developed the Database of Rheumatoid Arthritis-related Polymorphisms (RADB), to integrate all the RA-related genetic polymorphisms and provide a useful resource for researchers. We manually extracted the RA-related polymorphisms from 686 published reports, including RA susceptibility loci, polymorphisms associated with particular clinical features of RA, polymorphisms associated with drug response in RA and polymorphisms associated with a higher risk of cardiovascular disease in RA. Currently, RADB V1.0 contains 3235 polymorphisms that are associated with 636 genes and refer to 68 countries. The detailed information extracted from the literature includes basic information about the articles (e.g. PubMed ID, title and abstract), population information (e.g. country, geographic area and sample size) and polymorphism information (e.g. polymorphism name, gene, genotype, odds ratio and 95% confidence interval, P-value and risk allele). Meanwhile, useful annotations, such as hyperlinks to dbSNP, GenBank, UCSC, Gene Ontology and
Kyoto Encyclopedia of Genes and Genomes pathway, are included. In addition, a tool for meta-analysis was developed to summarize the results of multiple studies. The database is freely available at http://www.bioapp.org/RADB. 

**Database URL:** http://www.bioapp.org/RADB.

---

**Introduction**

Rheumatoid arthritis (RA) is a systemic inflammatory autoimmune disorder affected by genetic and environmental factors (1). The genetic component of RA has been estimated to be between 30 and 60% (2). Unlike single-gene disorders, RA is believed to be associated with multiple genes and their interactions (3). The strongest association has been shown to be with the HLA-DRB1 region (6p21), explaining ~30% of the total genetic effect (4). In addition to the HLA region, non-HLA genes (e.g. PTPN22, PADI4) have also been reported to contribute to RA susceptibility (5, 6). Currently, many loci that have convincing evidence for association with RA have been identified. However, the results are often poorly replicated, especially in different populations, increasing the complexity of the research. Collecting and collating the information about RA risk loci will facilitate systematic exploration of the genetic mechanisms of RA. Currently, there are several genetic association databases [e.g. Online Mendelian Inheritance in Man (OMIM) (7) and the Genetic Association Database (GAD) (8)] to store disease susceptibility loci. OMIM focuses on high-quality data of high significance for Mendelian disorders. Although in recent years, non-Mendelian diseases (also known as ‘common’ or ‘complex’ diseases) have been included, some biases still exist because of its history. In addition, OMIM is largely based on text and is a narrative history of disease research; thus, it is not designed to compare or analyze large sets of genetic data. More importantly, association studies of non-Mendelian diseases often have low-significance values, and findings of lower significance or negative findings are not routinely included in OMIM. Although GAD overcomes some disadvantages of OMIM, it is not a specialized database for RA, and polymorphisms associated with RA are not collected comprehensively. In addition, polymorphism genotype data are not collected in GAD, making some studies (e.g. meta-analysis) difficult. Therefore, a comprehensive, exhaustive and specialized database that includes all available genetic association study data from the published literature is urgently needed.

In addition to RA susceptibility, its clinical features [e.g. rheumatoid factor (RF) status, age of onset], drug response and cardiovascular (CV) events are also significantly influenced by genetic variation. Integrated management of these genetic variations and their relevant experimental information is also necessary, but so far, there is no database in which to store them.

Here, we present the Database of Rheumatoid Arthritis-related Polymorphisms (RADB) to integrate and analyze RA-related genetic polymorphisms extracted from published papers. The information collected comprises susceptibility loci for RA, polymorphisms associated with the clinical features of RA, polymorphisms associated with drug response in RA and polymorphisms associated with a higher risk of CV disease in RA. We not only collected polymorphisms that are significantly associated with RA, but also collected polymorphisms of lower significance and non-associated polymorphisms from RA-related research.

To facilitate the users’ ability to summarize the results of multiple studies, a linked tool for meta-analysis was developed. In addition, useful annotations, such as those from dbSNP (9), the National Centre for Biotechnology Information (NCBI) GenBank (10), University of California Santa Cruz (UCSC) (11) and Gene Ontology (GO) (12), were integrated into RADB to complement and extend the information from the literature.

---

**Data collection and database content**

**Data collection**

We searched the PubMed database with following keywords: ((polymorphism [Title/Abstract] OR polymorphisms [Title/Abstract] OR GWAS [Title/Abstract] OR GWA [Title/Abstract]) AND rheumatoid arthritis [Title/Abstract]) NOT review [Publication Type]. We obtained ~2000 publications. After manually scanning the list, 686 studies were included in RADB, comprising 21 candidate gene linkage analysis studies, 640 candidate gene association studies and 25 genome-wide association studies (GWAS). We extracted the important information from these reports, including basic information about the article [e.g. PubMed ID (PMID), title and abstract], population information (e.g. country, geographic area and sample size) and polymorphism information (e.g. polymorphism name, gene, genotype, odds ratio (OR) with 95% confidence interval (CI), P-value and risk allele).

Different laboratories may have different standards to describe the same polymorphism or gene; it is essential to
standardize them. Polymorphisms may have multiple names: for example, rs2476601, PTPN22 1858C/T and PTPN22 R620W represent the same polymorphism. To standardize the name, we merged the synonyms for each polymorphism. For genes, we used the approved gene name/symbol and Entrez Gene ID.

To obtain more information, we added hyperlinks to external databases: dbSNP or the IMGT/HLA database (13) for polymorphisms; and the NCBI Gene (14), EMBL-EBI (15), sequence databases (NCBI GenBank, RefSeq (16) and Unigene (17)), protein databases (Uniprot (18), PInt (19) and Prosite (20)) and biological pathway databases [GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (21)] for genes.

Data categories

Using our criteria, we identified 3235 polymorphisms from 636 genes. The polymorphisms were divided into four classes: (i) susceptibility loci for RA; (ii) polymorphisms associated with particular clinical features of RA; (iii) polymorphisms associated with drug response in RA; and (iv) polymorphisms associated with a higher risk of CV disease in RA. Although these four classes are not independent—for example, PTPN22 rs2476601 exists in all four classes—we believe that such classification will enable users to interrogate our database quickly and in more depth. The primary relationships between the classes are shown in Table 1.

(i) Susceptibility loci for RA

Currently, RADB contains 623 reports that examined the relationships between 2855 polymorphisms (597 genes/regions) and RA susceptibility. Among these, 562 polymorphisms (226 genes/regions) have \( P \) values < 0.05, 418 polymorphisms (180 genes/regions) have \( P \) values \( < 1 \times 10^{-3} \) and 242 polymorphisms (113 genes/regions) have \( P \) values \( < 1 \times 10^{-5} \). The strongest genetic association with RA has been found for HLA-DRB1 alleles on chromosome 6p21. In addition to the HLA region, non-HLA gene polymorphisms, including PTPN22 rs2476601 (5, 22), STAT4 rs7574865 (23, 24), TRAF1/C5 rs3761847 (25, 26), CTLA4 rs3087243 (27, 28) and PADI4 rs2240340 (6, 29), have also been reported to be strongly associated with RA susceptibility. However, these risk alleles differ among ethnic populations. HLA-DRB1*0401, 0404 and *0101 are the most common RA risk alleles among those of European ancestry (30, 31), while HLA-DRB1*0405 is the most common RA susceptibility allele for East Asian populations (32, 33). PTPN22 rs2476601 is a susceptibility locus for people of European ancestry, but is not associated with RA in Asian populations (6, 34, 35). Although an association has been reported between PADI4 rs2240340 and RA in East Asian populations, it was not replicated in those of European ancestry (36, 37). It was important, therefore, for our database to contain population information. The genes and genetic regions that have the strongest association with RA susceptibility are shown in Supplementary File S1 on the Web site: http://www.bioapp.org/research/RA.

(ii) Polymorphisms associated with the clinical features of RA

Currently, RADB contains 46 reports that examined the relationships between 156 polymorphisms (55 genes/regions) and clinical features of RA. Among these, 19 polymorphisms (18 genes/regions) have \( P \) values < 0.05 and 11 polymorphisms (5 genes/regions) have \( P \) values \( < 1 \times 10^{-3} \). The main clinical features analyzed include anti-citrullinated peptide antibodies (ACCP) status, RF status, age of onset and the activity/severity of RA. For example, HLA-DRB1 SE-alleles not only affect disease susceptibility, but also influence RF status, ACCP status, age of onset and the activity/severity of RA (38–42). In non-HLA regions, IL10 rs1800896 (–1082G/A) is associated with RF status (43). PTPN22 rs2476601 is associated with ACCP status (38). IL8 rs2227306 (781G/T) is associated with age of onset (44). RA activity/severity is influenced by IL6 rs1800795 (~174G/C), IL2 –330G/T and TNFA rs1800629 (~308A/G) (45–47).

(iii) Polymorphisms associated with drug response in RA

Disease-modifying anti-rheumatic drugs [e.g. methotrexate (MTX)] and biologics [e.g. anti-tumor necrosis factor (anti-TNF) agents] are the mainstay of treatment for RA. However, inconsistent response to these drugs is often observed, with considerable variability in both efficacy and toxicity (48). Currently, RADB contains 31 reports that examined the relationships between 176 polymorphisms (11 genes/regions) and drug response in RA. Among these, 40 polymorphisms (7 genes/regions) have \( P \) values < 0.05, and 13 polymorphisms (4 genes/regions) have \( P \) values

Table 1. Main relationships between the classes

<table>
<thead>
<tr>
<th>Relationship</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I ( \cap ) Class II</td>
<td>299</td>
</tr>
<tr>
<td>Class I ( \cap ) Class III</td>
<td>73</td>
</tr>
<tr>
<td>Class II ( \cap ) Class III</td>
<td>37</td>
</tr>
<tr>
<td>Class I ( \cap ) Class II ( \cap ) Class III</td>
<td>34</td>
</tr>
</tbody>
</table>

Class I represents susceptibility loci, Class II represents polymorphisms associated with clinical features, Class III represents polymorphisms associated with drug response, \( \cap \) represents intersection. The number of polymorphisms associated with a higher risk of cardiovascular events is not reflected in this table because there are insufficient reports (only 24).
<1 \times 10^{-3}. For example, RFC G80A, ATIC rs4673993, SHMT1 C1420T, SLC19A1 rs1232027 (G80A), HLA-DRB1, MTHFR rs1801133 (677CT) and MTHFR rs1801131 (1298AC) have been found to be associated with response to MTX treatment in patients with RA (49–54). The response to anti-TNF agents has been described to be associated with TNFA rs1800629 (−308GA), FCGR3A rs396991 (F158V), AFF3 rs10865035 and CD226 rs763361 (Gly307Ser) (55–59). The toxicity of MTX treatment has still relatively small, we expect the amount of data to expand.

(iv) Polymorphisms associated with a higher risk of CV disease in RA
RA is associated with an increased risk of CV events, causing increased CV morbidity and mortality (61). Currently, RADB contains 48 reports that examined the relationships between 83 polymorphisms (37 genes/regions) and a higher risk of CV in RA. Among these, 18 polymorphisms (17 genes/regions) have P values < 0.05, and 2 polymorphisms (2 genes/regions) have P values < 1 \times 10^{-3}, namely, LCE3C_LCE3B-del and CCR5 d32 (62–64). Although the number of polymorphisms associated with CV events is still relatively small, we expect the amount of data to expand on further research.

Meta-analysis module
The results of different association studies often show inconsistencies. A comprehensive evaluation of these results is important. Thus, we developed a module to perform a direct meta-analysis on the polymorphisms in RADB. Users can choose the parameters, such as the type of study (e.g. case–control study), the assumed risk allele and the genetic model. In addition, users can either analyze just their own data or supplement it with RADB data. In our meta-analysis module, the OR and 95% CI are calculated to assess the strength of association. Statistical heterogeneity among the studies is assessed with Woolf’s test (65). A fixed-effects model using the Mantel–Haenszel method (66) and the random effects model of DerSimonian and Laird (67) are used to summarize the results. The summary results are presented in tabular form and forest plots. We also provide a funnel plot to detect publication biases. The full paper hyperlinks of the included research are offered to facilitate the inquiries of users that want more detailed information of samples.

Querying the database
To meet the needs of different users, we offer different ways to search our database, including searching by polymorphism, searching by gene, searching by population, searching by different types of research (including candidate gene linkage analysis studies, candidate gene association studies and GWAS) and searching by chromosome.

Searching RADB by polymorphism name is a basic function. There are several types of polymorphism, such as single nucleotide polymorphisms, HLA alleles and micro-satelites. Users can use the dbSNP ‘rs’ number, gene symbol plus mutation position or gene symbol plus type of mutation to query RADB: for instance ‘rs2476601’, ‘PTPN22 1858CT’, ‘HLA-DRB1*0401’ or IL1RN 86 bp VNTR’ (Figure 1a). As mentioned above, the data are divided into four classes. Users can optionally choose a category of interest at this step. To facilitate ease of use, an auto-complete function has been used. The query results are reference centered (i.e. each record is a reference) and are displayed by publication date on a new page (Figure 1d). For instance, if a polymorphism has been described in 10 references there will be 10 records. The query results include basic information about the articles (e.g. PMID, title, source and important results/conclusions), population information (e.g. geographic area, population, population details and sample description) and polymorphism information (e.g. polymorphism name, gene symbol, Entrez Gene ID, genotype, OR and 95% CI, P-value and risk allele). If an article also examined other polymorphisms, a button will appear at the bottom of each record; users can click this button to display the other polymorphisms studied in the same paper.

Users can query the database using a keyword gene name (Fig. 1b) or list all the genes in RADB. Both Entrez Gene ID and Gene Symbol are currently supported, (e.g. 26191, PTPN22). The results are displayed on a new page (Fig. 1c). The results include gene-related information (e.g. number of references, number of polymorphisms and polymorphism list) and hyperlinked gene annotations (e.g. gene name, location, Entrez Gene, EMBL-EBI, UCSC, GenBank, RefSeq, Unigene, Uniprot, Pfam, Prosite, GO and KEGG pathway).

In addition to querying RADB by polymorphism name and gene name, users can search RADB by population, type of research and chromosome. If the user queries RADB by population, the results will list all studies undertaken within the same population and their corresponding polymorphisms. If the user searches RADB by type of research, the results will list all reports of the same study type and their corresponding polymorphisms. If the user searches RADB by chromosome (such as ‘6’, ‘X’ or ‘mitochondrion’), the results will list all the genes and their corresponding polymorphisms located in the queried chromosome or chromosomal region.
Submitting new data

To continually improve our database, we welcome the ongoing submission of new data. The submission process is simple. Users are only required to submit the article’s PMID and the corresponding polymorphism names. We will verify and input the data, if they meet our requirements, as soon as possible by manually filtering and sending data.

Discussion and conclusion

Over the past 4 years, we have extracted a large number of polymorphisms associated with RA from the published literature. These polymorphisms were collected and collated manually to obtain detailed and reliable data. The polymorphisms are associated with different phenotypes in different studies. For example, the purpose of some studies is to determine whether a certain polymorphism is an RA susceptibility locus; thus, we need to examine the association between the polymorphism and the presence of RA. In these cases, the samples are patients with RA and healthy controls. However, the purpose of other studies is to determine whether a certain polymorphism is associated with a particular clinical feature of RA, such as positivity for RF. The samples presented here would be RF+ and RF− patients. Our four data classifications make it convenient for researchers to access and query RADB for a specific purpose.

To obtain all the studies from a certain population, and to compare data for the same polymorphism in different populations, we collected population information that includes detailed geographical information, and we provide a corresponding method of query. Currently, RADB contains data from 68 countries (see Supplementary File S2 on the
Web site: http://www.bioapp.org/research/RA); however, only eight populations comprise >70% of the studies: Spain (96 studies), China (92 studies), UK (64 studies), Korea (55 studies), Japan (52 studies), USA (31 studies), Holland (31 studies), Sweden (24 studies), Poland (23 studies) and France (23 studies). Patients with RA are found worldwide, and the prevalence has been estimated at ~1% (2). Interestingly, the prevalence is higher (>2%) in some Native American populations, and is lower (<0.3%) in East Asian, Southeast Asian and African populations (68). More research on different populations will be beneficial to the understanding of the different genetic mechanisms involved in RA.

Table 2. Top 40 most significant GO terms associated with RA.

<table>
<thead>
<tr>
<th>GOID</th>
<th>TERM</th>
<th>Annotated</th>
<th>Significant</th>
<th>Expected</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0006955</td>
<td>Immune response</td>
<td>690</td>
<td>111</td>
<td>17.79</td>
<td>4.69E-58</td>
</tr>
<tr>
<td>GO:0002376</td>
<td>Immune system process</td>
<td>998</td>
<td>130</td>
<td>25.73</td>
<td>5.26E-58</td>
</tr>
<tr>
<td>GO:0002682</td>
<td>Regulation of immune system process</td>
<td>385</td>
<td>73</td>
<td>9.93</td>
<td>5.48E-42</td>
</tr>
<tr>
<td>GO:0048583</td>
<td>Regulation of response to stimulus</td>
<td>465</td>
<td>75</td>
<td>11.99</td>
<td>3.85E-38</td>
</tr>
<tr>
<td>GO:005776</td>
<td>Regulation of immune response</td>
<td>226</td>
<td>52</td>
<td>5.83</td>
<td>7.16E-34</td>
</tr>
<tr>
<td>GO:0002684</td>
<td>Positive regulation of immune system process</td>
<td>238</td>
<td>52</td>
<td>6.14</td>
<td>1.06E-32</td>
</tr>
<tr>
<td>GO:005896</td>
<td>Response to stimulus</td>
<td>3502</td>
<td>194</td>
<td>90.30</td>
<td>6.22E-32</td>
</tr>
<tr>
<td>GO:006952</td>
<td>Defense response</td>
<td>615</td>
<td>73</td>
<td>15.86</td>
<td>2.68E-28</td>
</tr>
<tr>
<td>GO:005865</td>
<td>Regulation of cell activation</td>
<td>175</td>
<td>41</td>
<td>4.51</td>
<td>6.38E-27</td>
</tr>
<tr>
<td>GO:0051239</td>
<td>Regulation of multicellular organismal process</td>
<td>937</td>
<td>87</td>
<td>24.16</td>
<td>2.87E-26</td>
</tr>
<tr>
<td>GO:002694</td>
<td>Regulation of leukocyte activation</td>
<td>166</td>
<td>39</td>
<td>4.28</td>
<td>1.21E-25</td>
</tr>
<tr>
<td>GO:006954</td>
<td>Inflammatory response</td>
<td>325</td>
<td>50</td>
<td>8.38</td>
<td>3.72E-24</td>
</tr>
<tr>
<td>GO:0042221</td>
<td>Response to chemical stimulus</td>
<td>1281</td>
<td>99</td>
<td>33.03</td>
<td>5.39E-24</td>
</tr>
<tr>
<td>GO:0048584</td>
<td>Positive regulation of response to stimulus</td>
<td>236</td>
<td>43</td>
<td>6.09</td>
<td>1.13E-23</td>
</tr>
<tr>
<td>GO:0001817</td>
<td>Regulation of cytokine production</td>
<td>181</td>
<td>38</td>
<td>4.67</td>
<td>3.78E-23</td>
</tr>
<tr>
<td>GO:0051249</td>
<td>Regulation of lymphocyte activation</td>
<td>148</td>
<td>35</td>
<td>3.82</td>
<td>4.24E-23</td>
</tr>
<tr>
<td>GO:0031347</td>
<td>Regulation of defense response</td>
<td>143</td>
<td>34</td>
<td>3.69</td>
<td>1.62E-22</td>
</tr>
<tr>
<td>GO:0050863</td>
<td>Regulation of T cell activation</td>
<td>117</td>
<td>31</td>
<td>3.02</td>
<td>5.35E-22</td>
</tr>
<tr>
<td>GO:0050778</td>
<td>Positive regulation of immune response</td>
<td>145</td>
<td>33</td>
<td>3.74</td>
<td>3.13E-21</td>
</tr>
<tr>
<td>GO:0006950</td>
<td>Response to stress</td>
<td>1685</td>
<td>110</td>
<td>43.45</td>
<td>5.19E-21</td>
</tr>
<tr>
<td>GO:0048518</td>
<td>Positive regulation of biological process</td>
<td>2033</td>
<td>123</td>
<td>52.42</td>
<td>5.31E-21</td>
</tr>
<tr>
<td>GO:0051704</td>
<td>Multiorganism process</td>
<td>681</td>
<td>66</td>
<td>17.56</td>
<td>1.20E-20</td>
</tr>
<tr>
<td>GO:0002697</td>
<td>Regulation of immune effector process</td>
<td>101</td>
<td>28</td>
<td>2.60</td>
<td>1.97E-20</td>
</tr>
<tr>
<td>GO:0080134</td>
<td>Regulation of response to stress</td>
<td>274</td>
<td>42</td>
<td>7.07</td>
<td>3.28E-20</td>
</tr>
<tr>
<td>GO:0051240</td>
<td>Positive regulation of multicellular organismal process</td>
<td>244</td>
<td>39</td>
<td>6.29</td>
<td>2.06E-19</td>
</tr>
<tr>
<td>GO:0010033</td>
<td>Response to organic substance</td>
<td>721</td>
<td>66</td>
<td>18.59</td>
<td>2.39E-19</td>
</tr>
<tr>
<td>GO:0002237</td>
<td>Response to molecule of bacterial origin</td>
<td>86</td>
<td>25</td>
<td>2.22</td>
<td>8.96E-19</td>
</tr>
<tr>
<td>GO:0009605</td>
<td>Response to external stimulus</td>
<td>914</td>
<td>74</td>
<td>23.57</td>
<td>1.00E-18</td>
</tr>
<tr>
<td>GO:0009607</td>
<td>Response to biotic stimulus</td>
<td>384</td>
<td>47</td>
<td>9.90</td>
<td>1.38E-18</td>
</tr>
<tr>
<td>GO:0050867</td>
<td>Positive regulation of cell activation</td>
<td>111</td>
<td>27</td>
<td>2.86</td>
<td>3.73E-18</td>
</tr>
<tr>
<td>GO:0019882</td>
<td>Antigen processing and presentation</td>
<td>83</td>
<td>24</td>
<td>2.14</td>
<td>5.74E-18</td>
</tr>
<tr>
<td>GO:0009611</td>
<td>Response to wounding</td>
<td>530</td>
<td>54</td>
<td>13.67</td>
<td>8.65E-18</td>
</tr>
<tr>
<td>GO:0001775</td>
<td>Cell activation</td>
<td>287</td>
<td>40</td>
<td>7.40</td>
<td>8.77E-18</td>
</tr>
<tr>
<td>GO:0002696</td>
<td>Positive regulation of leukocyte activation</td>
<td>106</td>
<td>26</td>
<td>2.73</td>
<td>1.40E-17</td>
</tr>
<tr>
<td>GO:0002819</td>
<td>Regulation of adaptive immune response</td>
<td>56</td>
<td>20</td>
<td>1.44</td>
<td>6.63E-17</td>
</tr>
<tr>
<td>GO:0050670</td>
<td>Regulation of lymphocyte proliferation</td>
<td>83</td>
<td>23</td>
<td>2.14</td>
<td>8.69E-17</td>
</tr>
<tr>
<td>GO:0048519</td>
<td>Negative regulation of biological process</td>
<td>1812</td>
<td>106</td>
<td>46.72</td>
<td>1.02E-16</td>
</tr>
<tr>
<td>GO:0070663</td>
<td>Regulation of leukocyte proliferation</td>
<td>84</td>
<td>23</td>
<td>2.17</td>
<td>1.15E-16</td>
</tr>
<tr>
<td>GO:0032944</td>
<td>Regulation of mononuclear cell proliferation</td>
<td>84</td>
<td>23</td>
<td>2.17</td>
<td>1.15E-16</td>
</tr>
<tr>
<td>GO:0051251</td>
<td>Positive regulation of lymphocyte activation</td>
<td>97</td>
<td>24</td>
<td>2.50</td>
<td>2.30E-16</td>
</tr>
</tbody>
</table>

Compared with analysis at the single-gene level, GO term enrichment analysis may provide further insight into the biological function of RA-related genes at the system level. GO term enrichment analysis for RA-related genes can be performed using Fisher’s exact test as implemented in the topGO package (69). A total of 477 genes (at least one polymorphism with a $P$-value < 0.05) have been associated with RA, and 364 of them have been successfully assigned GO terms. Table 2 lists the top 40 most significant GO terms (for more details, see Supplementary File S3 on the Web site: http://www.bioapp.org/research/RA), which include ‘inflammatory response’, ‘antigen processing and
presentation’ and ‘cytokine imbalances’; this is in agreement with a previous study (70). Over the past half century, several hypotheses have been proposed to explain the pathogenesis of RA. The key hypotheses are (i) the immune complex hypothesis and (ii) the T cells and cytokines hypothesis. The immune complex hypothesis states that immune complexes formed by antibodies and anti-antibodies (RFs) activate the complement cascade, which releases chemotactic factors such as C5a, resulting in inflammation and tissue damage (71). The T cells and cytokines hypothesis suggests that an imbalance between T helper 1 and T helper 2 cells and changes in cytokine expression (e.g. IL1, TNF-α and IL6) cause the immunopathological damage observed in RA (72). There is a close relationship between the enriched GO terms and these hypotheses. ‘Inflammatory response’ (GO:0006955 and GO:0006954) is a prominent characteristic of RA; ‘antigen processing and presentation’ (GO:0019882) is the initial step in the immune response (73). Moreover, ‘cytokine imbalances’ (GO:0001817) have been shown to be associated with many immunological processes, including promoting autoimmunity, chronic inflammation and tissue damage. Our results of GO term enrichment analysis show that the pathogenesis of RA is very complex. We suggest that more attention should be given to the enriched GO terms and the genes annotated to these categories.

RADB is a genetic database that has been developed for basic research and clinical application for RA. RADB has several advantages over OMIM and GAD. First, more detailed phenotypic data are provided in RADB. Second, RADB contains the genotype data of RA-related polymorphisms, which are not given in the other genetic databases. Third, meta-analysis can be directly performed in RADB. Last but not least, RADB offers an easy user interface and the data can be easily compared.

In the future, we intend to add proteomic and epigenetic information to RADB, to reflect the growing importance of mRNA expression, DNA methylation and microRNAs in the pathogenesis of RA (74–76). Because of its ability to integrate and analyze the data from different sources, we believe that RADB will be helpful in studying and identifying the genetic and molecular basis of RA.

**Supplementary data**

Supplementary data are available at Database Online.

**Funding**

This work was supported in part by grants from the National Natural Science Foundation of China (31200934, 81172842 and 81300945) and the Natural Science Foundation of Heilongjiang Province (grant numbers C201206 and QC2013063). Funding for open access charge: 31200934, 81172842 and C201206.

**Conflict of interest.** None declared.

**References**


