DATABASE
The Journal of Biological Databases and Curation

Database, 2014, 1–9
doi: 10.1093/database/bau090
Original article

Original article

RADB: a database of rheumatoid arthritis-related polymorphisms

Ruijie Zhang,[1,*,†], Meiwei Luan[1,†], Zhenwei Shang[1,†], Lian Duan[1,†] Guoping Tang[2,†], Miao Shi[1], Wenhua Lv[1], Hongjie Zhu[1], Jin Li[1], Hongchao Lv[1], Mingming Zhang[1], Guiyou Liu[3], He Chen[4,*] and Yongshuai Jiang[1,*]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China, [2]Yiwu Hospital, Zhejiang University, Yiwu 322000, China, [3]Genome Analysis Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China, [4]Depatment 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: zhangruijie 2013@gmail.com and He Chen. Tel: +86045186650721-106; Fax: +86045186615922; Email: chenhe201406@163.com

†These authors contributed equally to this work.

Citation details: Zhang,R., Luan,M., Shang,Z. *et al.*, RADB: a database of rheumatoid arthritis-related polymorphisms. *Database* (2014) Vol. 2014: article ID bau090; doi:10.1093/database/bau090

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 50 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), Pfam (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.

Table 1. Main relationships between the classes

Relationship	*n*
Class I ∩ Class II	299
Class I ∩ Class III	73
Class II ∩ Class III	37
Class I ∩ Class II ∩ Class III	34

Class I represents susceptibility loci, Class II represents polymorphisms associated with clinical features, Class III represents polymorphisms associated with drug response, ∩ 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).

(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 (781C/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

$< 1 \times 10^{-3}$. For example, *RFC* G80A, *ATIC* rs4673993, *SHMT1* C1420T, *SLC19A1* rs1232027 (G80A), *HLA-DRB1*, *MTHFR* rs1801133 (677C/T) and *MTHFR* rs1801131 (1298A/C) 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 (–308G/A), *FCGR3A* rs396991 (F158V), *AFF3* rs10865035 and *CD226* rs763361 (Gly307Ser) (55–59). The toxicity of MTX treatment has been shown to be associated with *RFC1* A80G, *MDR1* C3435T and *MTHFR* rs1801131 (60).

(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 microsatellites. 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* 1858C/T', '*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.

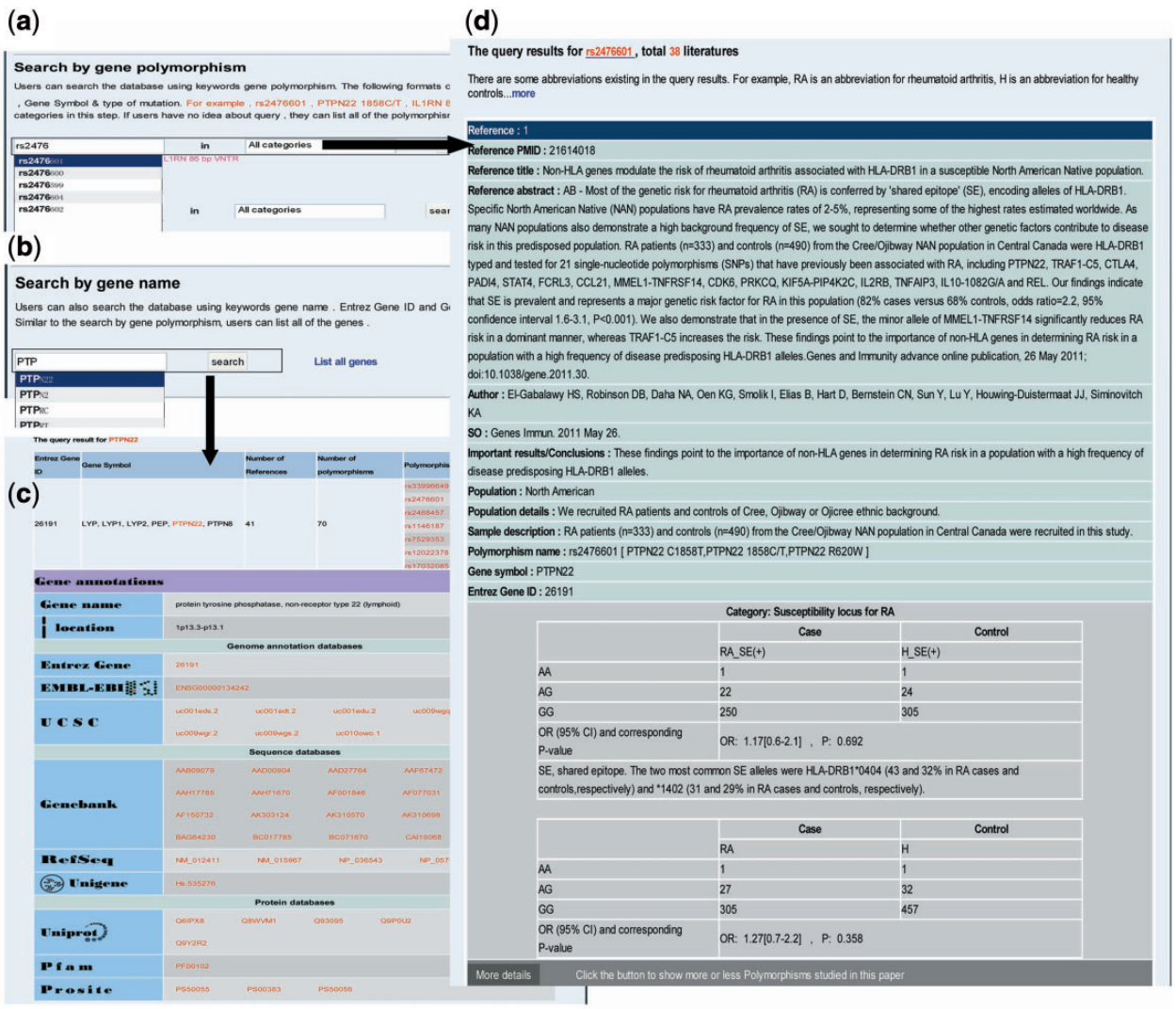

Figure 1. Examples of searching RADB by polymorphism name and gene name. (**a**) Searching RADB by polymorphism name. (**b**) Searching RADB by gene name. (**c**) Query results retrieved by searching with gene name. (**d**) Query results retrieved by searching with polymorphism name.

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.

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

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

GOID	TERM	Annotated	Significant	Expected	*P*-value
GO:0006955	Immune response	690	111	17.79	4.69E-58
GO:0002376	Immune system process	998	130	25.73	5.26E-58
GO:0002682	Regulation of immune system process	385	73	9.93	5.48E-42
GO:0048583	Regulation of response to stimulus	465	75	11.99	3.85E-38
GO:0050776	Regulation of immune response	226	52	5.83	7.16E-34
GO:0002684	Positive regulation of immune system process	238	52	6.14	1.06E-32
GO:0050896	Response to stimulus	3502	194	90.30	6.22E-32
GO:0006952	Defense response	615	73	15.86	2.68E-28
GO:0050865	Regulation of cell activation	175	41	4.51	6.38E-27
GO:0051239	Regulation of multicellular organismal process	937	87	24.16	2.87E-26
GO:0002694	Regulation of leukocyte activation	166	39	4.28	1.21E-25
GO:0006954	Inflammatory response	325	50	8.38	3.72E-24
GO:0042221	Response to chemical stimulus	1281	99	33.03	5.39E-24
GO:0048584	Positive regulation of response to stimulus	236	43	6.09	1.13E-23
GO:0001817	Regulation of cytokine production	181	38	4.67	3.78E-23
GO:0051249	Regulation of lymphocyte activation	148	35	3.82	4.24E-23
GO:0031347	Regulation of defense response	143	34	3.69	1.62E-22
GO:0050863	Regulation of T cell activation	117	31	3.02	5.35E-22
GO:0050778	Positive regulation of immune response	145	33	3.74	3.13E-21
GO:0006950	Response to stress	1685	110	43.45	5.19E-21
GO:0048518	Positive regulation of biological process	2033	123	52.42	5.31E-21
GO:0051704	Multiorganism process	681	66	17.56	1.20E-20
GO:0002697	Regulation of immune effector process	101	28	2.60	1.97E-20
GO:0080134	Regulation of response to stress	274	42	7.07	3.28E-20
GO:0051240	Positive regulation of multicellular organismal process	244	39	6.29	2.06E-19
GO:0010033	Response to organic substance	721	66	18.59	2.39E-19
GO:0002237	Response to molecule of bacterial origin	86	25	2.22	8.96E-19
GO:0009605	Response to external stimulus	914	74	23.57	1.00E-18
GO:0009607	Response to biotic stimulus	384	47	9.90	1.38E-18
GO:0050867	Positive regulation of cell activation	111	27	2.86	3.73E-18
GO:0019882	Antigen processing and presentation	83	24	2.14	5.74E-18
GO:0009611	Response to wounding	530	54	13.67	8.65E-18
GO:0001775	Cell activation	287	40	7.40	8.77E-18
GO:0002696	Positive regulation of leukocyte activation	106	26	2.73	1.40E-17
GO:0002819	Regulation of adaptive immune response	56	20	1.44	6.63E-17
GO:0050670	Regulation of lymphocyte proliferation	83	23	2.14	8.69E-17
GO:0048519	Negative regulation of biological process	1812	106	46.72	1.02E-16
GO:0070663	Regulation of leukocyte proliferation	84	23	2.17	1.15E-16
GO:0032944	Regulation of mononuclear cell proliferation	84	23	2.17	1.15E-16
GO:0051251	Positive regulation of lymphocyte activation	97	24	2.50	2.50E-16

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, 61300116, 81172842 and 81300945) and the Natural Science Foundation of Heilongjiang Province (grant numbers C201206 and QC2013C063). Funding for open access charge: 31200934, 81172842 and C201206.

Conflict of interest. None declared.

References

1. Barton,A., Thomson,W., Ke,X. *et al.* (2008) Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Hum. Mol. Genet.*, 17, 2274–2279.
2. Kurreeman,F.A., Padyukov,L., Marques,R.B. *et al.* (2007) A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med.*, 4, e278.
3. Deshmukh,H.A., Maiti,A.K., Kim-Howard,X.R. *et al.* (2011) Evaluation of 19 autoimmune disease-associated loci with rheumatoid arthritis in a colombian population: evidence for replication and gene-gene interaction. *J. Rheumatol.*, 38, 1866–1870.
4. Harrison,P., Pointon,J.J., Farrar,C., Harin,A. *et al.* (2007) MHC2TA promoter polymorphism (-168*G/A, rs3087456) is not associated with susceptibility to rheumatoid arthritis in British Caucasian rheumatoid arthritis patients. *Rheumatology (Oxford)*, 46, 409–411.
5. Begovich,A.B., Carlton,V.E., Honigberg,L.A. *et al.* (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am. J. Hum. Genet.*, 75, 330–337.
6. Lee,H.S., Korman,B.D., Le,J.M. *et al.* (2009) Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations. *Arthritis Rheum.*, 60, 364–371.
7. Hamosh,A., Scott,A.F., Amberger,J. *et al.* (2002) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, 30, 52–55.
8. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, 36, 431-432.
9. Sherry,S.T., Ward,M.H., Kholodov,M. *et al.* (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, 29, 308–311.
10. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J. *et al.* (2011) GenBank. *Nucleic Acids Res.*, 39, D32–D37.
11. Kuhn,R.M., Karolchik,D., Zweig,A.S. *et al.* (2009) The UCSC genome browser database: update *2009*. *Nucleic Acids Res.*, 37, D755–D761.
12. Ashburner,M., Ball,C.A., Blake,J.A. *et al.* (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29.
13. Robinson,J., Mistry,K., McWilliam,H. *et al.* (2011) The IMGT/HLA database. *Nucleic Acids Res.*, 39, D1171–D1176.
14. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, 33, D54–D58.
15. Hubbard,T.J., Aken,B.L., Beal,K. *et al.* (2007) Ensembl *2007*. *Nucleic Acids Res.*, 35, D610–D617.
16. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence

database of genomes, transcripts and proteins. *Nucleic Acids Res.*, 35, D61–D65.
17. Wheeler,D.L., Barrett,T., Benson,D.A. *et al.* (2007) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, 35, D5–D12.
18. UniProt Consortium. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.*, 39, D214–D219.
19. Finn,R.D., Mistry,J., Schuster-Böckler,B. *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.*, 34, D247–D251.
20. Hulo,N., Bairoch,A., Bulliard,V. *et al.* (2006) The PROSITE database. *Nucleic Acids Res.*, 34, D227–D230.
21. Kanehisa,M., Goto,S., Hattori,M. *et al.* (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, 34, D354–D357.
22. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447, 661–678.
23. Kobayashi,S., Ikari,K., Kaneko,H. *et al.* (2008) Association of STAT4 with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in the Japanese population. *Arthritis Rheum.*, 58, 1940–1946.
24. Remmers, E.F., Plenge,R.M., Lee,A.T. *et al.* (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.*, 357, 977–986.
25. Plenge,R.M., Seielstad,M., Padyukov,L. *et al.* (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. *N. Engl. J. Med.*, 357, 1199–1209.
26. Zhu,J., Zhang,D., Wu,F. *et al.* (2011) Single nucleotide polymorphisms at the TRAF1/C5 locus are associated with rheumatoid arthritis in a Han Chinese population. *BMC Med. Genet.*, 12, 53.
27. Walker,E.J., Hirschfield,G.M., Xu,C. *et al.* (2009) CTLA4/ICOS gene variants and haplotypes are associated with rheumatoid arthritis and primary biliary cirrhosis in the Canadian population. *Arthritis Rheum.*, 60, 931–937.
28. Danoy,P., Wei,M., Johanna,H. *et al.* (2011) Association of variants in MMEL1 and CTLA4 with rheumatoid arthritis in the Han Chinese population. *Ann. Rheum. Dis.*, 70, 1793–1797.
29. Ikari,K., Kuwahara,M., Nakamura,T. *et al.* (2005) Association between PADI4 and rheumatoid arthritis: a replication study. *Arthritis Rheum.*, 52, 3054–3057.
30. Lopez-Arbesu,R., Ballina-García,F.J., Alperi-López,M. *et al.* (2007) MHC class I chain-related gene B (MICB) is associated with rheumatoid arthritis susceptibility. *Rheumatology (Oxford)*, 46, 426–430.
31. Hajeer,A.H., Dababneh,A., Makki,R.F. *et al.* (2000) Different gene loci within the HLA-DR and TNF regions are independently associated with susceptibility and severity in Spanish rheumatoid arthritis patients. *Tissue Antigens*, 55, 319–325.
32. Ichikawa,N., Kotake,S., Hakoda,M. *et al.* (2009) Combining effects of polymorphism of tumor necrosis factor alpha 5′-flanking region and HLA-DRB1 on radiological progression in patients with rheumatoid arthritis. *Mod. Rheumatol.*, 19, 134–139.
33. Kochi,Y., Yamada,R., Kobayashi,K. *et al.* (2004) Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences. *Arthritis Rheum.*, 50, 63–71.
34. Chabchoub,G., Teixiera,E.P., Maalej,A. *et al.* (2009) The R620W polymorphism of the protein tyrosine phosphatase 22 gene in autoimmune thyroid diseases and rheumatoid arthritis in the Tunisian population. *Ann. Hum. Biol.*, 36, 342–349.
35. Ikari,K., Momohara,S., Inoue,E. *et al.* (2006) Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. *Rheumatology (Oxford)*, 45, 1345–1348.
36. Burr,M.L., Naseem,H., Hinks,A. *et al.* (2010) PADI4 genotype is not associated with rheumatoid arthritis in a large UK Caucasian population. *Ann. Rheum. Dis.*, 69, 666–670.
37. Martinez,A., Valdivia,A., Pascual-Salcedo,D. *et al.* (2005) PADI4 polymorphisms are not associated with rheumatoid arthritis in the Spanish population. *Rheumatology (Oxford)*, 44, 1263–1266.
38. Szodoray,P., Szabó,Z., Kapitány,A. *et al.* (2010) Anti-citrullinated protein/peptide autoantibodies in association with genetic and environmental factors as indicators of disease outcome in rheumatoid arthritis. *Autoimmun. Rev.*, 9, 140–143.
39. Mattey,D.L., Dawes,P.T., Clarke,S. *et al.* (2002) Relationship among the HLA-DRB1 shared epitope, smoking, and rheumatoid factor production in rheumatoid arthritis. *Arthritis Rheum.*, 47, 403–407.
40. van Gaalen,F.A., van Aken,J., Huizinga,T.W. *et al.* (2004) Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis. *Arthritis Rheum.*, 50, 2113–2121.
41. Okada,Y., Yamada,R., Suzuki,A. *et al.* (2009) Contribution of a haplotype in the HLA region to anti-cyclic citrullinated peptide antibody positivity in rheumatoid arthritis, independently of HLA-DRB1. *Arthritis Rheum.*, 60, 3582–3590.
42. Turesson,C. and Matteson,E.L. (2006) Genetics of rheumatoid arthritis. *Mayo Clin. Proc.*, 81, 94–101.
43. Nemec,P., Goldbergova,M.P., Gatterova,J. *et al.* (2009) Association of polymorphisms in interleukin-10 gene promoter with autoantibody production in patients with rheumatoid arthritis. *Ann. N. Y. Acad. Sci.*, 1173, 501–508.
44. Emonts,M., Hazes,M.J., Houwing-Duistermaat,J.J. *et al.* (2011) Polymorphisms in genes controlling inflammation and tissue repair in rheumatoid arthritis: a case control study. *BMC Med. Genet.*, 12, 36.
45. Pawlik,A., Wrzesniewska,J., Florczak,M. *et al.* (2005) IL-6 promoter polymorphism in patients with rheumatoid arthritis. *Scand. J. Rheumatol.*, 34, 109–113.
46. Pawlik,A., Kurzawski,M., Florczak,M. *et al.* (2005) IL1beta+3953 exon 5 and IL-2 -330 promoter polymorphisms in patients with rheumatoid arthritis. *Clin. Exp. Rheumatol.*, 23, 159–164.
47. Hussein,Y.M., Mohamed,R.H., Pasha,H.F. *et al.* (2011) Association of tumor necrosis factor alpha and its receptor polymorphisms with rheumatoid arthritis in female patients. *Cell Immunol.*, 271, 192–196.
48. Ranganathan,P. (2008) Pharmacogenomics in rheumatoid arthritis. *Methods Mol. Biol.*, 448, 413–435.
49. Hayashi,H. Fujimaki,C., Daimon,T. *et al.* (2009) Genetic polymorphisms in folate pathway enzymes as a possible marker for predicting the outcome of methotrexate therapy in Japanese

patients with rheumatoid arthritis. *J. Clin. Pharm. Ther.*, 34, 355–361.
50. Drozdzik,M., Rudas,T., Pawlik,A. *et al.* (2007) Reduced folate carrier-1 80G>A polymorphism affects methotrexate treatment outcome in rheumatoid arthritis. *Pharmacogenomics J.*, 7, 404–407.
51. Lee,Y.C. Cui,J., Costenbader,K.H. *et al.* (2009) Investigation of candidate polymorphisms and disease activity in rheumatoid arthritis patients on methotrexate. *Rheumatology (Oxford)*, 48, 613–617.
52. James,H.M., Gillis,D., Hissaria,P. *et al.* (2008) Common polymorphisms in the folate pathway predict efficacy of combination regimens containing methotrexate and sulfasalazine in early rheumatoid arthritis. *J. Rheumatol.*, 35, 562–571.
53. Ali,A.A., Moatter,T., Baig,J.A. *et al.* (2006) Polymorphism of HLA-DR and HLA-DQ in rheumatoid arthritis patients and clinical response to methotrexate—a hospital-based study. *J. Pak. Med. Assoc.*, 56, 452–456.
54. Kurzawski,M., Pawlik,A., Safranow,K. *et al.* (2007) 677C>T and 1298A>C MTHFR polymorphisms affect methotrexate treatment outcome in rheumatoid arthritis. *Pharmacogenomics*, 8, 1551–1559.
55. Maxwell,J.R., Potter,C., Hyrich,K.L. *et al.* (2008) Association of the tumour necrosis factor-308 variant with differential response to anti-TNF agents in the treatment of rheumatoid arthritis. *Hum. Mol. Genet.*, 17, 3532–3538.
56. Mugnier,B., Balandraud,N., Darque,A. *et al.* (2003) Polymorphism at position -308 of the tumor necrosis factor alpha gene influences outcome of infliximab therapy in rheumatoid arthritis. *Arthritis Rheum.*, 48, 1849–1852.
57. Canete,J.D., Suárez,B., Hernández,M.V. *et al.* (2009) Influence of variants of Fc gamma receptors IIA and IIIA on the American College of Rheumatology and European League Against Rheumatism responses to anti-tumour necrosis factor alpha therapy in rheumatoid arthritis. *Ann. Rheum. Dis.*, 68, 1547–1552.
58. Morales-Lara,M.J., Conesa-Zamora,P., García-Simón,M.S. *et al.* (2010) Association between the FCGR3A V158F polymorphism and the clinical response to infliximab in rheumatoid arthritis and spondyloarthritis patients. *Scand. J. Rheumatol.*, 39, 518–520.
59. Tan,R.J., Gibbons,L.J., Potter,C. *et al.* (2010) Investigation of rheumatoid arthritis susceptibility genes identifies association of AFF3 and CD226 variants with response to anti-tumour necrosis factor treatment. *Ann. Rheum. Dis.*, 69, 1029–1035.
60. Bohanec Grabar,P., Logar,D., Lestan,B. and Dolzan,V. (2008) Genetic determinants of methotrexate toxicity in rheumatoid arthritis patients: a study of polymorphisms affecting methotrexate transport and folate metabolism. *Eur. J. Clin. Pharmacol.*, 64, 1057–1068.
61. Quyyumi,A.A. (2006) Inflamed joints and stiff arteries: is rheumatoid arthritis a cardiovascular risk factor? *Circulation*, 114, 1137–1139.
62. Palomino-Morales,R., Gonzalez-Juanatey,C., Vazquez-Rodriguez,T.R. *et al.* (2010) A1298C polymorphism in the MTHFR gene predisposes to cardiovascular risk in rheumatoid arthritis. *Arthritis Res. Ther.*, 12, R71.
63. Panoulas,V.F., Stavropoulos-Kalinoglou,A., Metsios,G.S. *et al.* (2009) Association of interleukin-6 (IL-6)-174G/C gene polymorphism with cardiovascular disease in patients with rheumatoid arthritis: the role of obesity and smoking. *Atherosclerosis*, 204, 178–183.
64. Teruel,M., Martin,J.E., González-Juanatey,C. *et al.* (2011) Association of acid phosphatase locus 1*C allele with the risk of cardiovascular events in rheumatoid arthritis patients. *Arthritis Res. Ther.*, 13, R116.
65. Woolf,B. (1955) On estimating the relation between blood group and disease. *Ann. Hum. Genet.*, 19, 251–253.
66. Mantel,N. and Haenszel,W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.*, 22, 719–748.
67. DerSimonian,R. and Laird,N. (1986) Meta-analysis in clinical trials. *Control. Clin. Trials*, 7, 177–188.
68. Carmona,L., Villaverde,V., Hernández-García,C. *et al.* (2002) The prevalence of rheumatoid arthritis in the general population of Spain. *Rheumatology (Oxford)*, 41, 88–95.
69. Gentleman,R.C., Carey,V.J. and Bates,D.M. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, 5, R80.
70. Ballard,D.H., Aporntewan,C., Lee,J.Y. *et al.* (2009) A pathway analysis applied to Genetic Analysis Workshop 16 genome-wide rheumatoid arthritis data. *BMC Proc.*, 3(Suppl. 7), S91.
71. Zvaifler,N.J. (1973) The immunopathology of joint inflammation in rheumatoid arthritis. *Adv. Immunol.*, 16, 265–336.
72. McInnes,I.B. and Schett,G. (2007) Cytokines in the pathogenesis of rheumatoid arthritis. *Nat. Rev. Immunol.*, 7, 429–442.
73. Aarvak,T. and Natvig,J.B. (2001) Cell-cell interactions in synovitis: antigen presenting cells and T cell interaction in rheumatoid arthritis. *Arthritis Res.*, 3, 13–17.
74. Batliwalla,F.M., Baechler,E.C., Xiao,X. *et al.* (2005) Peripheral blood gene expression profiling in rheumatoid arthritis. *Genes Immun.*, 6, 388–397.
75. Richardson,B., Scheinbart,L., Strahler,J. *et al.* (1990) Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Rheum.*, 33, 1665–1673.
76. Stanczyk,J., Pedrioli,D.M., Brentano,F. *et al.* (2008) Altered expression of MicroRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis. *Arthritis Rheum.*, 58, 1001–1009.