Published online 16 October 2022 *Nucleic Acids Research, 2023, Vol. 51, Database issue* ***D1381–D1387***
https://doi.org/10.1093/nar/gkac850

RABC: Rheumatoid Arthritis Bioinformatics Center

Haiyan Chen[1,2,†], Jing Xu[1,2,†], Siyu Wei[1,2,†], Zhe Jia[1,2,†], Chen Sun[1,2,†], Jingxuan Kang[1,2], Xuying Guo[1], Nan Zhang[1], Junxian Tao[1,2], Yu Dong[1], Chen Zhang[1,2], Yingnan Ma[1], Wenhua Lv[1], Hongsheng Tian[1], Shuo Bi[1], Hongchao Lv[1], Chen Huang[3,4], Fanwu Kong[5], Guoping Tang[6,*], Yongshuai Jiang[1,2,*] and Mingming Zhang[1,2,*]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, [2]The ABC Project, Harbin, China, [3]Dr. Neher's Biophysics Laboratory for Innovative Drug Discovery, Macau University of Science and Technology, Macau, SAR, China, [4]Stat Key laboratory of Quality Research in Chinese Medicine, Macau Institute For Applied Research in Medicine and Health, Macau University of Science and Technology, Macau, SAR, China, [5]Department of Nephrology, The Second Affiliated Hospital, Harbin Medical University, Harbin, China and [6]The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu, China

Received August 14, 2022; Revised September 13, 2022; Editorial Decision September 15, 2022; Accepted September 21, 2022

ABSTRACT

Advances in sequencing technologies have led to the rapid growth of multi-omics data on rheumatoid arthritis (RA). However, a comprehensive database that systematically collects and classifies the scattered data is still lacking. Here, we developed the Rheumatoid Arthritis Bioinformatics Center (RABC, http://www.onethird-lab.com/RABC/), the first multi-omics data resource platform (data hub) for RA. There are four categories of data in RABC: (i) 175 multi-omics sample sets covering transcriptome, epigenome, genome, and proteome; (ii) 175 209 differentially expressed genes (DEGs), 105 differentially expressed microRNAs (DEMs), 18 464 differentially DNA methylated (DNAm) genes, 1 764 KEGG pathways, 30 488 GO terms, 74 334 SNPs, 242 779 eQTLs, 105 m6A-SNPs and 18 491 669 meta-mQTLs; (iii) prior knowledge on seven types of RA molecular markers from nine public and credible databases; (iv) 127 073 literature information from PubMed (from 1972 to March 2022). RABC provides a user-friendly interface for browsing, searching and downloading these data. In addition, a visualization module also supports users to generate graphs of analysis results by inputting personalized parameters. We believe that RABC will become a valuable resource and make a significant contribution to the study of RA.

INTRODUCTION

Rheumatoid arthritis (RA) is a classic autoimmune disease that affects ∼0.24% to 1% of the population worldwide (1). The inflammatory response in RA patients causes pain, swelling, stiffness and loss of function in the joints, which can lead to severe disability if not treated promptly (2–4). Accordingly, RA causes serious inconvenience to patients as well as imposes a great burden on society (5). With the development of sequencing technology, massive data of RA have emerged, involving transcriptome, epigenome, genome, proteome and so on (6). These data can evaluate the repeatability of current data analysis methods and offer substantial help to solve diverse RA-related biological questions, including identifying the biomarkers and exploring the pathogenesis (7–10). Each type of data provides a valuable resource for RA research, but they are scattered in different databases with different operational requirements. In addition, standardized methods for the same type of raw data are not uniform, which will affect the merging and analysis of data at late stage (11). Thus, it is essential to use a standardized pipeline to pre-process these data and manage it in a findable, accessible, interoperable, and reusable manner (12). Nevertheless, a comprehensive user-friendly database that could solve the above problems and offer some analysis results and visualization functions meanwhile is still lacking.

RABC, a comprehensive RA bioinformatics center was developed here to fill the gap. Dispersed RA data were collected and classified, and multi-omics data were processed using a unified standardized pipeline. RABC stores these data and allows users to browse, search and download the content of interest, in which the visualization

*To whom correspondence should be addressed. Tel: +86 451 86620941; Fax: +86 451 86620941; Email: zhangmingming@hrbmu.edu.cn
Correspondence may also be addressed to Yongshuai Jiang. Tel: +86 451 86620941; Fax: +86 451 86620941; Email: jiangyongshuai@gmail.com
Correspondence may also be addressed to Guoping Tang. Email: tgp@zju.edu.cn.
†The authors wish it to be known that, in their opinion, the first five authors should be regarded as Joint First Authors.

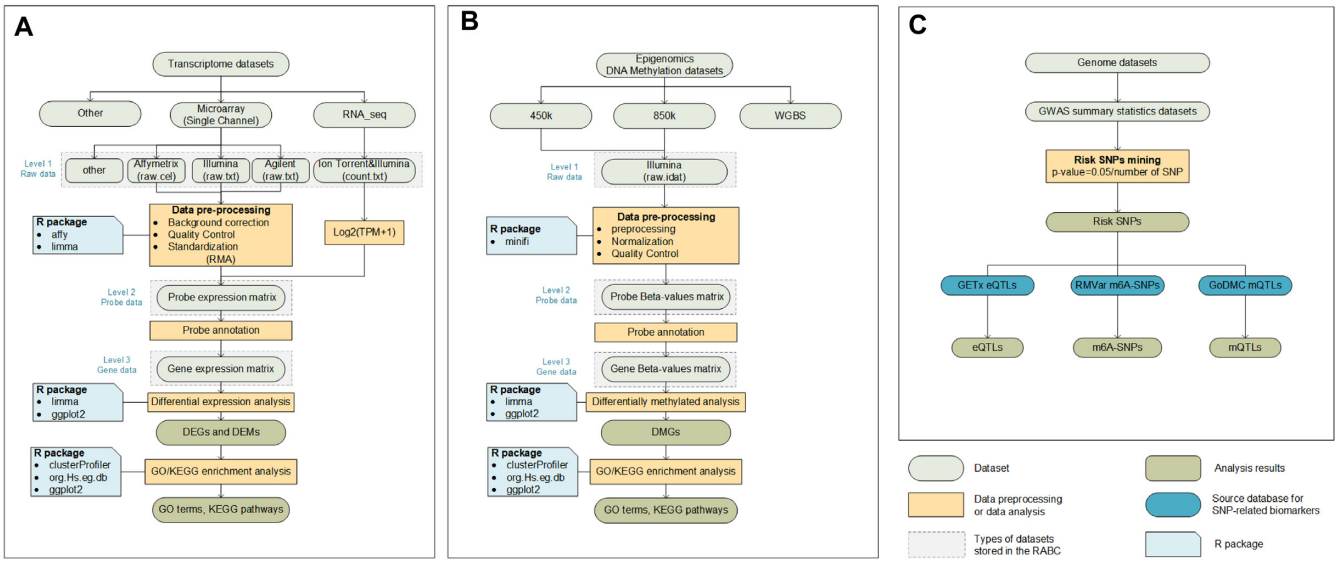

Figure 1. Multi-omics data processing pipeline. (**A**) Transcriptome data processing pipeline; (**B**) DNA methylation data processing pipeline; (**C**) GWAS summary statistics dataset processing pipeline.

module supports users, even those without a bioinformatics background, to visualize the analysis results using RABC data. In summary, RABC is by far the most comprehensive and systematic RA database that addresses the current main limitations of storing and using large amounts of RA data. We envision that RABC will provide valuable RA-related data sources, thereby accelerating scientific discovery around the world and making new contributions to RA research.

MATERIALS AND METHODS

Multi-omics data collection and preprocessing

For transcriptome, raw data were downloaded from GEO (13), SRA (14) and ArrayExpress (15) databases, mainly including microarray chip data of mRNA, lncRNA, microRNA and RNA-sequencing (RNA-seq) data. Microarray raw data from different platforms were preprocessed using the unified RMA normalization method (16,17) of different R packages. Expression values of RNA-seq datasets were normalized as $\log_2(\text{TPM} + 1)$ values. The probes were re-annotated to gene symbols according to the corresponding platform annotation file. The pipeline of transcriptome data was shown in Figure 1A.

For epigenome, DNA methylation raw data of the Illumina HumanMethylation450 BeadChip (450K), Infinium MethylationEPIC BeadChip (850K), and the bed files of whole-genome bisulfite sequencing (WGBS) data were downloaded from the GEO, SRA and EWAS project (18,19). Methylation array datasets from 450K and 850K were normalized uniformly using the PreprecessIllumina function in the minfi package (20) through the Bioconductor project (21) and DNA methylation values for each locus in each sample were described as beta values (22) (Supplementary File M5). Lastly, probes were re-annotated to gene symbols based on the Illumina HumanMethylation450/850 BeadChip annotation information. The pipeline of epigenome data was shown in Figure 1B.

For genome, eight GWAS summary statistics datasets of RA-healthy were obtained from the GWAS catalog (23), Wellcome Trust Case Control Consortium (WTCCC), and Genetic Analysis Workshop 16 (GAW16) (Supplementary File Table S2). The pipeline of genome data was shown in Figure 1C.

For proteome, only five datasets were obtained from GEO. Due to the small amount of data, the current version only provides raw data and does not perform data processing and analysis.

Some datasets have multiple grouped samples for different research purposes, but to better compare the differences between RA samples and samples in other states, we prefer to extract double-grouped samples for analysis in a single dataset. So, the preprocessed multi-group datasets were reintegrated (Supplementary File M6). After data reintegration, each RABC datasets contained two groups or only RA samples. For the double-grouped datasets, RA samples were defined as a case group (group1) and one of another status (healthy samples, other disease samples, or samples related to drug treatment) were defined as a control group (group2). These the double-grouped datasets can be used for differential analysis (Table S3). An example was GSE1919, which contains three groups of samples (RA, osteoarthritis (OA), and healthy). It was kept RA as group 1 and reintegrated into two subsets: RA (case)—healthy (control) and RA (case)—OA (control) (Table S4).

Finally, after the data collection and preprocessing process (Supplementary File M2-4, Table S1), the downloaded raw datasets, normalized matrix data that facilitate reanalysis, gene expression/methylation profiling datasets, and GWAS summary statistics datasets are stored in RABC (Supplementary Figure S3).

RA-related biomarkers and biology functions

The differentially expressed genes/microRNAs (DEGs/DEMs) and differentially DNA methylation genes (DMGs) were identified by the R package 'limma' (using lmFit and eBayes function). In order to further understand the biological functions of DEGs and DMGs, GO (24) terms and KEGG (25) pathways were identified using the clusterProfiler package (26).

RA-related risk SNPs were identified from 8 GWAS datasets. To further explore the regulatory mechanisms of these risk SNPs, they were integrated with GTEx (27), RMVar (28), and GoDMC (29) database to identify risk eQTLs, risk m6A-SNPs and risk mQTL. Finally, the results of these analyses were also included in the RABC.

The detail analysis process and code were shown in Supplementary File M7, M10.

RA-related prior knowledge of biomarkers

To offer more comprehensive information on RA-related biomarkers, RABC gathered the RA-related biomarkers that were previously identified and scattered in various databases focusing on human diseases and disease-related molecular markers: PedAM (30), Circ2Disease (31), MiRNASNP-v3 (32), HMDD (33), LncRNADisease (34), DisGeNet (35), CTD (36), LncRNADisease v2.0 (37) and DiseaseMeth 2.0 (38). The information on RA-related biomarkers was firstly disposaled and combined by searching the databases with the keyword 'Rheumatoid Arthritis or RA', and then was divided into seven parts according to the types of biomarkers, including genes, SNPs, lncRNAs, microRNAs, circRNAs, pre-miRNAs and DNA methylation genes. They provide a wealth of prior knowledge for RA research. The detail of these biomarkers was shown in the RABC.

RA-related literature

RABC database aims to provide users with a 'one-stop' understanding of RA-related content as much as possible. Therefore, information on RA-related literature (from 1972 to March 2022) was searched from PubMed by the keyword 'Rheumatoid arthritis/RA', and was saved in PubMed format. Then, the key information, including PMID, abstract, title, publication date and keywords was extracted from PubMed files (Supplementary File M9).

Database implementation

Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), JavaScript and Hypertext Preprocessor (PHP) were used to build the web interface. Tens of millions of records were stored in the open-source MySQL database. The service of RABC was deployed in the Apache web server. Data analyses were mainly carried out using the R and Perl (Table S5).

DATABASE CONTENT

RABC provides researchers with rich, high-quality RA datasets and information, including multi-omics sample datasets, biomarkers and biology functions, RA-related prior knowledge, and literature information (Supplementary Figure S1). Among them, the multi-omics raw data has been processed through a unified standardized process. All content is available in the web interface (Figure 2).

RABC data statistics

Overall, RABC is freely available at http://www.onethird-lab.com/RABC/. It contains four categories of data (Table 1).

In the multi-omics dataset, 175 multi-omics datasets were collected, covering ∼900 000 samples. There are 140 datasets for transcriptome, 22 datasets for epigenome, 8 case–control GWAS summary statistics datasets for genome and 5 protein datasets for proteome.

In the RA-related biomarkers and biology functions, the standardized preprocessed multi-omics data were analyzed to obtain 175 209 DEGs (Supplementary Figure S4), 105 DEMs, 18 464 DMGs (Supplementary Figure S8), 1764 KEGG pathways, 30 491 GO terms, 74 334 risk SNPs, 242 779 eQTLs from 49 tissues, 105 m6A-SNPs and 18 491 669 meta-mQTLs.

In the RA-related prior knowledge, 887 SNPs, 152 microRNAs, 20 lncRNAs, 32 271 genes, 2528 DNA methylation genes 20 circRNAs and 22 pre-miRNAs were obtained from nine databases.

In literature, 127 073 publications in PubMed from 1972 to March 2022 were extracted. The gradually increasing research on RA in recent years (Supplementary Figure S2) shows that the construction of RABC is necessary and feasible.

Web interface

RABC provides a user-friendly web interface that enables users to browse, search, and download the above-mentioned four categories of data. It also provides a visual analysis result module that supports users to customize the visual graphics of interest.

Users can quickly browse the data of interest through the home page (Supplementary Figure S10) or browse modules (Supplementary Figure S11). First, in the multi-omics data browsing interface (Supplementary Figure S12), users can view the description information of the data. The detailed information interface of each dataset (Supplementary Figure S13) provides not only a more detailed data description but also the raw data, preprocessed matrix data and analysis results. Browse modules also support viewing the results of differential analysis and gene expression volcano plots for gene expression data. Second, on the RA-related biomarkers and biology functions page, users can quickly browse DEGs (Supplementary Figure S14), DMGs, DEMs, GO terms, KEGG pathways, risk SNPs, risk eQTLs, risk m6A-SNPs and risk mQTLs. RA-related prior knowledge (Supplementary Figures S15 and S16) and relevant information of RA-related literature (Supplementary Figure S17) can also be browsed and searched in the RABC browsing interface (Supplementary File E1).

The search interface supports users to query the content of RABC through a variety of conditions, including dataset ID, data type, gene name, PMID, etc (Supplementary File

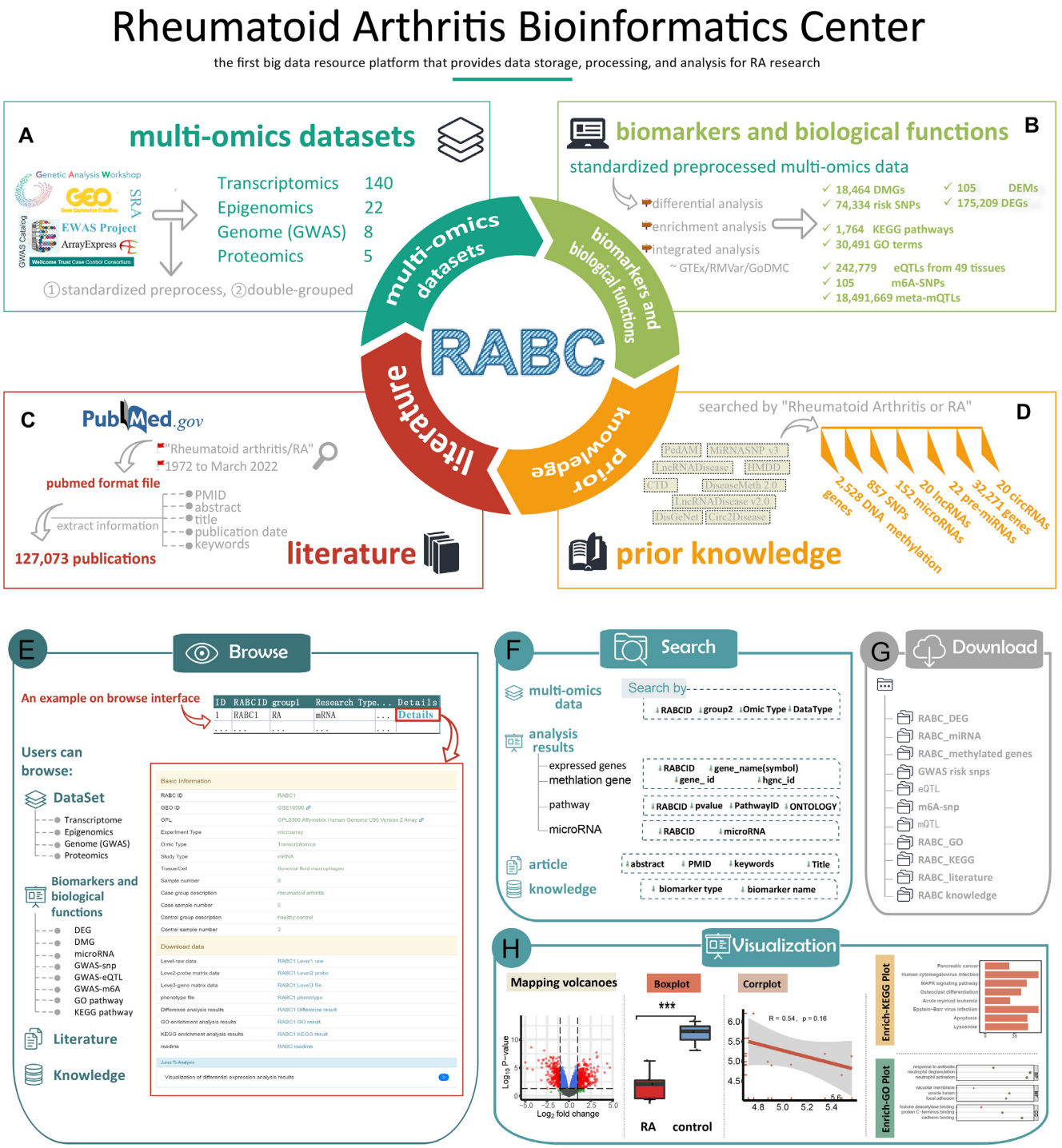

Figure 2. Schematic view of RABC. (**A**) multi-omics datasets; (**B**) RA-related biomarkers and biology functions; (**C**) RA-related literature; (**D**) RA-related prior knowledge; (**E**) browse module; (**F**) search module; (**G**) download module; (**H**) visualization module.

E2, Figures S18 and S19). Users can directly download biomarkers and biology functions files and prior knowledge from the download module (Supplementary File E3, Figure S20).

The visualization module supports drawing volcano plots of differential expression analysis, boxplots of gene expression, scatter plots of correlation between two genes, bar charts, and bubble charts of functional enrichment analysis for gene expression profiling dataset (Supplementary Figures S5–S7). These visualization functions only require the user to select the dataset and input the specified threshold or parameters, and then the corresponding graph can be drawn. These graphs also can be supported download (Supplementary File E4, Figure S21).

Table 1. Data statistics of RABC

Data	Number
Datasets:	175
transcriptome	140
epigenome	22
genome	8
proteome	5
Samples	≈900 000
DEGs	175 209
MicroRNAs	105
DMGs	18 464
KEGG pathways	1764
GO terms:	30 491
Biological Process (BP)	24 635
Cellular Component (CC)	3925
Molecular Function (MF)	1931
GWAS risk SNPs	74 334
GTEx tissue types	49
Risk eQTLs	242 779
Risk m6A-SNPs	105
Risk mQTLs	18 491 669
Database of prior knowledge sources	9
Prior knowledge:	35 870
Risk SNPs	857
microRNA	152
lncRNA	20
genes	32 271
DNA methylation genes	2528
circRNA	20
pre-miRNAs	22
Literature (1972–2022.3)	127 073

A case of RA analysis using the RABC dataset

In several independent studies, the Protein Tyrosine Phosphatase Non-Receptor Type 22 (*PTPN22*) gene has been shown to be associated with multiple autoimmune diseases (39–41). Gene expression meta-analysis can improve the reproducibility of analysis results by integrating multiple microarray datasets (42), and it has also been used in several studies to find disease-related risk biomarkers (43–45). To verify the consistency of the differential expression of the *PTPN22* gene in different studies, a meta-analysis was performed using 27 gene expression datasets from RABC (Table S6). The results of meta-analysis showed significant association between *PTPN22* and RA (P-value = 0.0024, Table S7). Following a search for '*PTPN22*' in the prior knowledge section, there are three prior evidences supporting the association between *PTPN22* and RA (Supplementary Figure S9). In addition, the searching results in the literature section also showed a total of 373 literature mentioned that *PTPN22* is related to RA. The case above suggests the potential of RABC in providing data and prior knowledge for RA research and the construction of RABC will greatly help the research of RA.

DISCUSSION AND FUTURE DEVELOPMENTS

RABC is the first big data resource platform that provides data storage, processing, and analysis for RA research. It not only solves the current problems in the use of RA data, but also brings more well-categorized and uniformly processed data, and multiple data analysis results. The practical and user-friendly platform provides researchers to explore biomarkers relevant to the pathogenesis, diagnosis, and treatment of RA. RABC is by far the largest and most comprehensive RA database. The emergence of this informative and easy-to-navigate database will advance the development of RA research and enhance the in-depth understanding of RA and even autoimmune diseases, thereby advancing the life sciences.

At present, immune system research is a hot topic. RA is the most common autoimmune disorder that can cause severe joint pain and is associated with other complex diseases. Therefore, RABC can be used in comparison with multiple diseases, such as various cancers, diabetes, heart disease, Alzheimer's disease, eye disease, joint disease and other autoimmune diseases.

As the cost of high-throughput sequencing continues to decline and more and more researchers focus on RA research, different high-throughput platforms will generate more and more RA sequencing data. In the future, we will continue to update the RABC and add more RA-related data, such as CHIP-seq and ATAC-seq. At the same time, single-cell omics data and spatial transcriptomics data will also be considered to be added in our subsequent updates. In addition, on the basis of a large amount of RA data, more effective analytical methods and tools will be developed for RA research in RABC.

DATA AVAILABILITY

All data is freely available from the RABC web interface. The code related to the data processing and analysis is freely available from GitHub (https://github.com/onethird-lab/RABC-Code).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China [31970651, 92046018, 81601422, 81600403]; Mathematical Tianyuan Fund of the National Natural Science Foundation of China [12026414]; Basic Public Welfare Research Project of Zhejiang Province [LGF20C060001]; Funding Project for Innovative Scientific Research-Special Scientific Research Project for Basic Medical Youth [2021-KYYWF-0280]; Science and Technology Development Fund, Macau SAR 462 [0020/2021/A, 001/2020/ALC, SKL-QRCM (MUST)-2020–2022]; Dr Neher's Biophysics Laboratory for Innovative Drug Discovery, State Key Laboratory of Quality Research in Chinese Medicine, Macau University of Science and Technology, Macau, China [FRG-21-032-SKL]. Funding for open access charge: National Natural Science Foundation of China; Mathematical Tianyuan Fund of the National Natural Science Foundation of China; Basic Public Welfare Research Project of Zhejiang Province; Funding Project for Innovative Scientific Research-Special Scientific Research Project for Basic Medical Youth.

Conflict of interest statement. None declared.

REFERENCES

1. Almutairi,K., Nossent,J., Preen,D., Keen,H. and Inderjeeth,C. (2021) The global prevalence of rheumatoid arthritis: a meta-analysis based on a systematic review. *Rheumatol. Int.*, **41**, 863–877.
2. Smolen,J.S., Aletaha,D. and McInnes,I.B. (2016) Rheumatoid arthritis. *Lancet*, **388**, 2023–2038.
3. Radu,A.-F. and Bungau,S.G. (2021) Management of rheumatoid arthritis: an overview. *Cells*, **10**, 2857.
4. Walsh,D.A. and McWilliams,D.F. (2014) Mechanisms, impact and management of pain in rheumatoid arthritis. *Nat. Rev. Rheumatol.*, **10**, 581–592.
5. Cross,M., Smith,E., Hoy,D., Carmona,L., Wolfe,F., Vos,T., Williams,B., Gabriel,S., Lassere,M., Johns,N. *et al.* (2014) The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. *Ann. Rheum. Dis.*, **73**, 1316–1322.
6. Bender,E. (2015) Big data in biomedicine. *Nature*, **527**, Cp4–Cp4.
7. Svendsen,A.J., Holm,N.V., Kyvik,K., Petersen,P.H. and Junker,P. (2002) Relative importance of genetic effects in rheumatoid arthritis: historical cohort study of danish nationwide twin population. *BMJ*, **324**, 264–266.
8. van der Woude,D., Houwing-Duistermaat,J.J., Toes,R.E., Huizinga,T.W., Thomson,W., Worthington,J., van der Helm-van Mil,A.H. and de Vries,R.R. (2009) Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum.*, **60**, 916–923.
9. Smith,S.L., Plant,D., Eyre,S. and Barton,A. (2013) The potential use of expression profiling: implications for predicting treatment response in rheumatoid arthritis. *Ann. Rheum. Dis.*, **72**, 1118–1124.
10. Guo,S., Xu,L., Chang,C., Zhang,R., Jin,Y. and He,D. (2020) Epigenetic regulation mediated by methylation in the pathogenesis and precision medicine of rheumatoid arthritis. *Front. Genet.*, **11**, 811.
11. Ristevski,B. and Chen,M. (2018) Big data analytics in medicine and healthcare. *J. Integr. Bioinformatics*, **15**, 20170030.
12. Deng,N., Wu,C., Yaseen,A. and Wu,H. (2022) ImmuneData: an integrated data discovery system for immunology data repositories. *Database: the Journal of Biological Databases and Curation*, **2022**, baac003.
13. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
14. Kodama,Y., Shumway,M., Leinonen,R. and International Nucleotide Sequence Database,C. (2012) The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–D56.
15. Athar,A., Fullgrabe,A., George,N., Iqbal,H., Huerta,L., Ali,A., Snow,C., Fonseca,N.A., Petryszak,R., Papatheodorou,I. *et al.* (2019) ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res.*, **47**, D711–D715.
16. Irizarry,R.A., Bolstad,B.M., Collin,F., Cope,L.M., Hobbs,B. and Speed,T.P. (2003) Summaries of affymetrix genechip probe level data. *Nucleic Acids Res.*, **31**, e15.
17. Lopez-Romero,P., Gonzalez,M.A., Callejas,S., Dopazo,A. and Irizarry,R.A. (2010) Processing of agilent microRNA array data. *BMC Res. Notes*, **3**, 18.
18. Xu,J., Zhao,L., Liu,D., Hu,S., Song,X., Li,J., Lv,H., Duan,L., Zhang,M., Jiang,Q. *et al.* (2018) EWAS: epigenome-wide association study software 2.0. *Bioinformatics*, **34**, 2657–2658.
19. Liu,D., Zhao,L., Wang,Z., Zhou,X., Fan,X., Li,Y., Xu,J., Hu,S., Niu,M., Song,X. *et al.* (2019) EWASdb: epigenome-wide association study database. *Nucleic Acids Res.*, **47**, D989–D993.
20. Aryee,M.J., Jaffe,A.E., Corrada-Bravo,H., Ladd-Acosta,C., Feinberg,A.P., Hansen,K.D. and Irizarry,R.A. (2014) Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. *Bioinformatics*, **30**, 1363–1369.
21. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
22. Houseman,E.A., Christensen,B.C., Yeh,R.F., Marsit,C.J., Karagas,M.R., Wrensch,M., Nelson,H.H., Wiemels,J., Zheng,S., Wiencke,J.K. *et al.* (2008) Model-based clustering of DNA methylation array data: a recursive-partitioning algorithm for high-dimensional data arising as a mixture of beta distributions. *BMC Bioinf.*, **9**, 365.
23. Buniello,A., MacArthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E. *et al.* (2019) The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
24. Gene Ontology Consortium (2021) The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res.*, **49**, D325–D334.
25. Kanehisa,M., Furumichi,M., Sato,Y., Ishiguro-Watanabe,M. and Tanabe,M. (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.*, **49**, D545–D551.
26. Yu,G., Wang,L.G., Han,Y. and He,Q.Y. (2012) clusterProfiler: an r package for comparing biological themes among gene clusters. *Omics*, **16**, 284–287.
27. GTEx Consortium (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
28. Luo,X., Li,H., Liang,J., Zhao,Q., Xie,Y., Ren,J. and Zuo,Z. (2021) RMVar: an updated database of functional variants involved in RNA modifications. *Nucleic Acids Res.*, **49**, D1405–D1412.
29. Min,J.L., Hemani,G., Hannon,E., Dekkers,K.F., Castillo-Fernandez,J., Luijk,R., Carnero-Montoro,E., Lawson,D.J., Burrows,K., Suderman,M. *et al.* (2021) Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.*, **53**, 1311–1321.
30. Jia,J., An,Z., Ming,Y., Guo,Y., Li,W., Li,X., Liang,Y., Guo,D., Tai,J., Chen,G. *et al.* (2018) PedAM: a database for pediatric disease annotation and medicine. *Nucleic Acids Res.*, **46**, D977–D983.
31. Yao,D., Zhang,L., Zheng,M., Sun,X., Lu,Y. and Liu,P. (2018) Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease. *Sci. Rep.*, **8**, 11018.
32. Liu,C.J., Fu,X., Xia,M., Zhang,Q., Gu,Z. and Guo,A.Y. (2021) miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. *Nucleic Acids Res.*, **49**, D1276–D1281.
33. Huang,Z., Shi,J., Gao,Y., Cui,C., Zhang,S., Li,J., Zhou,Y. and Cui,Q. (2019) HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.*, **47**, D1013–D1017.
34. Chen,G., Wang,Z., Wang,D., Qiu,C., Liu,M., Chen,X., Zhang,Q., Yan,G. and Cui,Q. (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.*, **41**, D983–D986.
35. Pinero,J., Ramirez-Anguita,J.M., Sauch-Pitarch,J., Ronzano,F., Centeno,E., Sanz,F. and Furlong,L.I. (2020) The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.*, **48**, D845–D855.
36. Grondin,C.J., Davis,A.P., Wiegers,J.A., Wiegers,T.C., Sciaky,D., Johnson,R.J. and Mattingly,C.J. (2021) Predicting molecular mechanisms, pathways, and health outcomes induced by juul e-cigarette aerosol chemicals using the comparative toxicogenomics database. *Curr. Res. Toxicol.*, **2**, 272–281.
37. Bao,Z., Yang,Z., Huang,Z., Zhou,Y., Cui,Q. and Dong,D. (2019) LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.*, **47**, D1034–D1037.
38. Xiong,Y., Wei,Y., Gu,Y., Zhang,S., Lyu,J., Zhang,B., Chen,C., Zhu,J., Wang,Y., Liu,H. *et al.* (2017) DiseaseMeth version 2.0: a major expansion and update of the human disease methylation database. *Nucleic Acids Res.*, **45**, D888–D895.
39. Tizaoui,K., Terrazzino,S., Cargnin,S., Lee,K.H., Gauckler,P., Li,H., Shin,J.I. and Kronbichler,A. (2021) The role of PTPN22 in the pathogenesis of autoimmune diseases: a comprehensive review. *Semin. Arthritis Rheum.*, **51**, 513–522.
40. Tizaoui,K., Shin,J.I., Jeong,G.H., Yang,J.W., Park,S., Kim,J.H., Hwang,S.Y., Park,S.J., Koyanagi,A. and Smith,L. (2022) Genetic polymorphism of PTPN22 in autoimmune diseases: a comprehensive review. *Medicina (Mex)*, **58**, 1034.
41. Pradhan,V., Borse,V. and Ghosh,K. (2010) PTPN22 gene polymorphisms in autoimmune diseases with special reference to systemic lupus erythematosus disease susceptibility. *J. Postgrad. Med.*, **56**, 239–242.
42. Sweeney,T.E., Haynes,W.A., Vallania,F., Ioannidis,J.P. and Khatri,P. (2017) Methods to increase reproducibility in differential gene expression via meta-analysis. *Nucleic Acids Res.*, **45**, e1.

43. Likhitrattanapisal,S., Tipanee,J. and Janvilisri,T. (2016) Meta-analysis of gene expression profiles identifies differential biomarkers for hepatocellular carcinoma and cholangiocarcinoma. *Tumour Biol.*, **37**, 12755–12766.
44. Makashir,S.B., Kottyan,L.C. and Weirauch,M.T. (2015) Meta-analysis of differential gene co-expression: application to lupus. *Pacific Symp. Biocomput.*, 443–454.
45. Bell,R., Barraclough,R. and Vasieva,O. (2017) Gene expression meta-analysis of potential metastatic breast cancer markers. *Curr. Mol. Med.*, **17**, 200–210.