

CVD Atlas: a multi-omics database of cardiovascular disease

Qiheng Qian^{1,2,3,†}, Ruikun Xue^{1,2,3,†}, Chenle Xu^{1,2,3}, Fengyu Wang⁴, Jingyao Zeng^{1,2} and Jingfa Xiao^{1,2,3,*}

¹National Genomics Data Center, China National Center for Bioinformation, Beijing 100101, China

²Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

³University of Chinese Academy of Sciences, Beijing 100049, China

⁴Department of Neurology, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou, Henan 450003, China

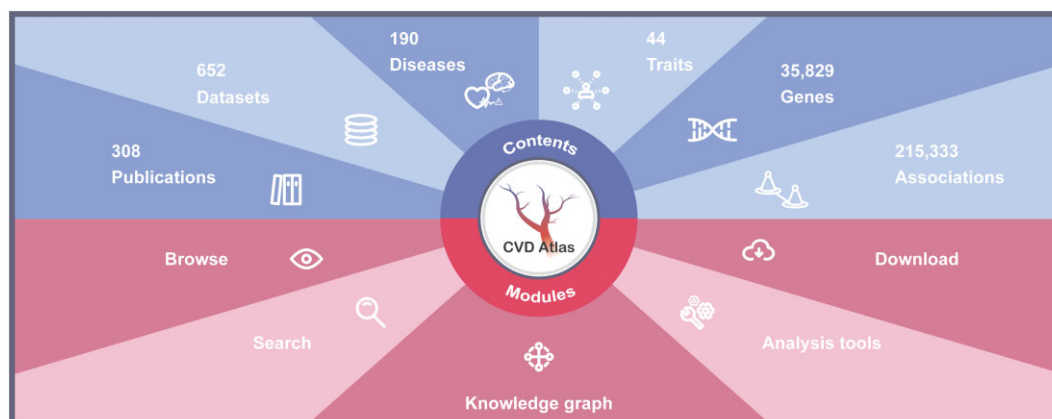
*To whom correspondence should be addressed. Tel: +86 10 8409 7443; Fax: +86 10 8409 7720; Email: xiaojingfa@big.ac.cn

[†]The first two authors should be regarded as Joint First Authors.

Abstract

Cardiovascular disease (CVD) is the leading cause of illness and death worldwide. Numerous studies have been conducted into the underlying mechanisms and molecular characteristics of CVD using various omics approaches. However, there is still a need for comprehensive resources on CVD. To fill this gap, we present the CVD Atlas, accessed at <https://ngdc.cncb.ac.cn/cvd>. This database compiles knowledge and information from manual curation, large-scale data analysis, and existing databases, utilizing multi-omics data to understand CVDs comprehensively. The current version of CVD Atlas contains 215,333 associations gathered from 308 publications, 652 datasets and 7 databases. It covers 190 diseases and 44 traits across multiple omics levels. Additionally, it provides an interactive knowledge graph that integrates disease-gene associations and two types of analysis tools, offering an engaging way to query and display relationships. CVD Atlas also features a user-friendly web interface that allows users to easily browse, search, and download all association information, research metadata, and annotation details. In conclusion, CVD Atlas is a valuable resource that enhances the accessibility and utility of knowledge and information related to CVD, benefiting human health and CVD research communities.

Graphical abstract



Introduction

Cardiovascular disease (CVD) is the primary cause of illness and death globally, resulting in 17.8 million deaths in 2017, a 21.1% increase from 2007 to 2017 (1). The number reached 20.5 million in 2021 (2). CVD includes a range of disorders affecting the heart and blood vessels, such as coronary artery disease (CAD) and stroke. Various omics studies on CVD have revealed some important genes associated with the disease.

For example, a comprehensive gene-centric association study identified four genetic variants linked to myocardial infarction (3). RNA-seq analysis and network construction pinpointed lncRNA *SNHG8* as a risk factor for acute myocardial infarction, supported by RT-PCR for its diagnostic value (4). In addition, methylation of the *PTEN* promoter was found to be involved in cerebral cavernous malformation (CCM), a type of vascular lesion in the central nervous system (5). A proteomic

Received: August 15, 2024. Revised: September 9, 2024. Editorial Decision: September 14, 2024. Accepted: September 18, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

study also suggested that *LRP2* and *SZT2* may play a role in CVD (6). An untargeted metabolomic study on coronary heart disease (CHD) identified four lipid-related metabolites with a causal role in CHD (7). These research findings, encompassing genomics, transcriptomics, epigenomics, proteomics, and metabolomics, offer a comprehensive understanding of the molecular mechanisms underlying CVD. Furthermore, significant research efforts have been made to identify CVD-related traits, such as blood lipid levels, blood pressure, and inflammatory markers, through genome-wide association studies (GWAS) to elucidate the genetic loci associated with these risk factors for CVD (8–10).

The advancement of sequencing technology has provided valuable insights into the mechanisms and molecular signatures of cardiovascular disease (CVD). As a result, there is now an abundance of CVD-related data and knowledge available. To better understand the biological characteristics of CVD, it is crucial to develop integrated and organized multi-omics databases. These databases are designed to provide a comprehensive view of the complex biological processes underlying CVD, including CaGE (11), CADgene (12), CardioGenBase (13) and C/VDdb (14). CaGE is a knowledgebase of cardiac gene expression, presenting 7,188 unique Human LocusLink identifiers expressed in human normal or dilated cardiomyopathic failing cardiac tissue. CADgene provides evidence related to coronary artery disease and myocardial infarction, with information on more than 300 candidate genes from over 1,300 publications of genetic studies. CardioGenBase uses a text mining approach to collect gene-disease associations, containing about 1,500 CVD genes and 24,000 research articles on six major CVDs. C/VDdb is an integrated multi-omics information resource of CVD, consisting of 4,353 unique molecule entries collected from nearly 100 manually curated studies. However, these existing databases for cardiovascular diseases (CVD) have some limitations. Firstly, these databases often only focus on a narrow range of CVD disorders, neglecting others requiring comprehensive characterization. Secondly, the abundant omics data urgently calls for unified analysis, improved integration, and visualization to unlock their full potential and gain comprehensive insights. Thirdly, prior knowledge from other published databases must be utilized more effectively. Lastly, continuously updating or making these databases readily available is crucial.

To tackle these issues, we have developed CVD Atlas, a comprehensive database for cardiovascular disease. CVD Atlas includes manually curated gene–disease associations, analysis results from human multi-omics datasets and prior knowledge from existing databases. Furthermore, it boasts an interactive, information-rich knowledge graph and provides two types of analysis tools. CVD Atlas offers the latest integrated multi-omics resources for cardiovascular disease research. Consequently, we believe that CVD Atlas will significantly support cardiovascular disease research and clinical practice.

Materials and methods

Data curation and integration

Our process began with a comprehensive literature search in NCBI PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). We used predefined keywords such as ‘(Cardiovascular disease) AND (GWAS)’, ‘(Cardiovascular disease) AND (Transcrip-

tome)’ and ‘(Cardiovascular disease) AND (Methylation)’. We then reviewed the articles rigorously, ensuring that only the most relevant and high-quality publications were included. Subsequently, we manually curated the study information from each relevant publication, a step that underscores the human expertise and judgment involved in the process. We selected only associations reported at significant levels (P -value < 0.05 or adjusted P -value < 0.05) for inclusion in CVD Atlas. To create a more comprehensive database, we also reviewed and integrated prior knowledge from existing databases such as PedAM (15), TWAS Atlas (16), HMDD v4.0 (17), miRNASNP-v3 (18), CTD (19), PharmGKB (20) and RNADisease v4.0 (21).

Multi-omics data collection and processing

We obtained GWAS summary statistics for cardiovascular disease (CVD) from the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) (22). We processed the raw data using MungeSumstats v1.10.1 (23) to ensure consistency with the reference genome GRCh38 (hg38). We conducted various quality control steps such as filling in missing SNP IDs, chromosome (CHR), base pair position (BP), and statistical values like effect size (Beta), standard error (SE), odds ratio (OR) and P -value using available data. We then used a P -value threshold of $5.0E-8$ to identify significant SNPs. If no SNPs reached the genome-wide significance level in a GWAS dataset, we defined a genetic locus using the suggestive P -value threshold of $1.0E-4$. For fine-mapping, we performed colocalization analysis using the R package coloc v5.2.2 (24), integrating GWAS and eQTL data from 49 tissues in GTEx v8 (<https://gtexportal.org/>) (25). We defined genetic loci as 1 MB windows around lead SNPs with the smallest P -value and used recommended thresholds ($PP4 \geq 0.75$ and $PP4/PP3 \geq 3$) to establish causal gene associations.

For transcriptomics, we collected microarray and RNA-sequencing (RNA-seq) data from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) (26,27). The gene expression profiles of the microarray data were obtained using the advanced R package GEOquery (28). We performed normalization and differential expression analysis using the powerful R package limma (29). The bulk RNA-seq data were downloaded, and we used the state-of-the-art tools fastp v0.20.0 (30), STAR v2.7.6a (31) and RSEM v1.3.1 (32) for quality control, read alignment, and estimation of gene expression levels, respectively. We utilized DESeq2 v1.38.3 (33) for differential expression analysis, identifying differentially expressed genes based on $|\log_2 FCI| \geq 1$ and adjusted P -value < 0.05 . GO enrichment and co-expression analyses were performed using the R packages clusterProfiler v3.18.0 (34) and WGCNA v1.70-3 (35), respectively.

We obtained beta-value matrices for methylation data from the Illumina HumanMethylation450 BeadChip (450K) from GEO. Then, we used the R package ChAMP v2.32.0 (36–38) to identify differentially methylated positions (DMPs). DMPs with $|\Delta \text{beta}| \geq 0.2$ and adjusted P -value < 0.05 were considered significant.

For proteomics, we downloaded protein expression profiles from PRIDE (<https://www.ebi.ac.uk/pride/>) (39) and conducted the differential expression analysis using the R package DEP v1.12.0 (40). We considered proteins with leverage $\log_2 FCI \geq 1$ and P -value < 0.05 as significantly differentially expressed proteins.

For metabolomics, we carefully selected metabolomic datasets from the Metabolomics Workbench (<https://www.metabolomicsworkbench.org/>) (41) and performed the differential analysis using the R package MetaboAnalystR v4.0.0 (42). This rigorous selection process ensures the high quality and reliability of our data. We identified differential metabolites with a variable importance projection (VIP) score ≥ 1 using Partial Least Squares Discriminant Analysis (PLS-DA).

Ontology mapping and standardizing

To standardize disease and trait names, we employed disease-related terms and identifiers from various ontologies, such as the Disease Ontology (DO, <https://disease-ontology.org/>), Experimental Factor Ontology (EFO, <https://www.ebi.ac.uk/efo/>), Medical Subject Headings (MeSH, <https://www.nlm.nih.gov/mesh/meshhome.html>), and MONDO (<https://mondo.monarchinitiative.org/>). Utilizing these ontologies, we organized diseases and traits into eight and five distinct subcategories, respectively.

Knowledge graph construction

We have recently built a knowledge graph to improve the visualization and understanding of connections. This graph identifies three main entities: disease, trait and gene. The graph consists of three main sections: ‘Disease network’, ‘Trait network’ and ‘Gene network’ each centered around its respective core entity. We calculated a confidence score for each disease-gene or trait-gene association based on multi-omics evidence to provide a deeper understanding. This score is included in the graph as an additional attribute, indicating the strength and reliability of each connection.

Enrichment analysis and signature comparison

The current version of CVD Atlas includes two tools: ‘Enrichment analysis’ and ‘Signature comparison’. The ‘Enrichment analysis’ tool determines gene function based on the gene-disease or gene-trait perspective, using all genes in the repository as a reference set. Users can submit a list of gene symbols, set conditions, and calculate the significance *P*-value using Fisher’s exact test. The ‘Signature comparison’ tool allows users to compare user-submitted signatures with differentially expressed genes from each dataset in CVD Atlas.

Database implementation

CVD Atlas was developed using SpringBoot (<https://spring.io/projects/spring-boot>) and MyBatis (<https://mybatis.org/mybatis-3>) as the back-end framework. The web interface was built using HTML (HyperText Markup Language), CSS (Cascading Style Sheets), jQuery (<https://jquery.com>), AJAX (Asynchronous JavaScript and XML) and Thymeleaf (<https://www.thymeleaf.org>). Front-end frameworks such as Semantic UI (<https://semantic-ui.com>) and Bootstrap (<https://getbootstrap.com>) were also employed. Data were stored and queried using MySQL (<https://www.mysql.com>), and visualization was achieved through Apache ECharts (<http://echarts.apache.org>).

Database contents and usage

CVD Atlas is a comprehensive database created through meticulous manual curation, analysis of multi-omics

datasets and integration of knowledge from existing databases. Additionally, the website includes a visual and interactive knowledge graph and two types of tools to facilitate a better understanding of the intricate relationships between genes and CVD (Figure 1A).

Integration of multi-omics association knowledge for cardiovascular disease

In the current version, CVD Atlas houses 215,333 associations from 308 publications, 652 datasets and 7 existing databases. These encompass 35,829 genes, 190 diseases grouped into 8 categories, and 44 traits classified into 5 categories (Figure 1B, C). Among these genes, the most prevalent are protein-coding genes, followed by lncRNA (Supplementary Figure S1). Most genes are linked to multiple diseases or traits, with a median association of four diseases or traits per gene. The top 20 genes with the highest number of associated diseases or traits exhibit various functions across different conditions (Figure 1D). *hsa-mir-146a* is the most prominent gene associated with 55 diseases, while *BRAP*, *HECTD4* and *NAA25* are the leading genes linked to 19 traits. Conversely, the median number of related genes for 190 diseases and 44 traits is 86 and 234, respectively. The top 10 diseases and traits with the most associated genes are shown in Figure 1E. Hypertension is associated with 836 high-quality genes with a confidence score greater than 0.7. This indicates its complex pathological mechanisms and molecular functions. *AGT*, *NOS3*, *POMC*, *ADRB1* and *NPPA* are the top genes associated with hypertension, ranked by their confidence scores. *AGT* gene encodes angiotensinogen, the sole precursor of all angiotensin peptides, playing a vital role in regulating blood pressure (43). The associations between *AGT* polymorphism and hypertension have been well-documented (44–46). *NOS3* encodes one of the three isoforms of nitric oxide synthase, primarily found in endothelial cells, and is essential for vascular health. *NOS3* is also recognized as a genetic risk factor for hypertension (47,48). Additionally, the CVD Atlas contains 457,286 SNPs, 8,436 differentially methylated positions, 453 differentially expressed proteins and 148 differentially expressed metabolites.

User-friendly modules for accessing data

The ‘Browse’ page of CVD Atlas contains seven interactive tables that are easy to navigate. Each table is indexed by disease, trait, dataset, gene, SNP, association and publication, allowing users to access relevant data easily. The browse tables provide basic information and brief summarized statistics. For example, the ‘Disease browse’ table displays all the diseases included in CVD Atlas, along with their names, CVD Atlas IDs and statistics such as significant SNPs, colocalized genes, differentially expressed genes, proteins, metabolites, differentially methylated positions, associated chemicals, other associations and datasets. Users can access detailed information on each item by visiting its dedicated page, which provides comprehensive information on all associations across different omics levels (Figure 2A).

To facilitate efficient querying, CVD Atlas offers several search channels: (i) a quick search box on the home page for real-time queries by specifying disease name, trait name, gene name, SNP, or CVD Atlas ID; (ii) an advanced search function on the ‘Search’ page that allows users to access CVD Atlas

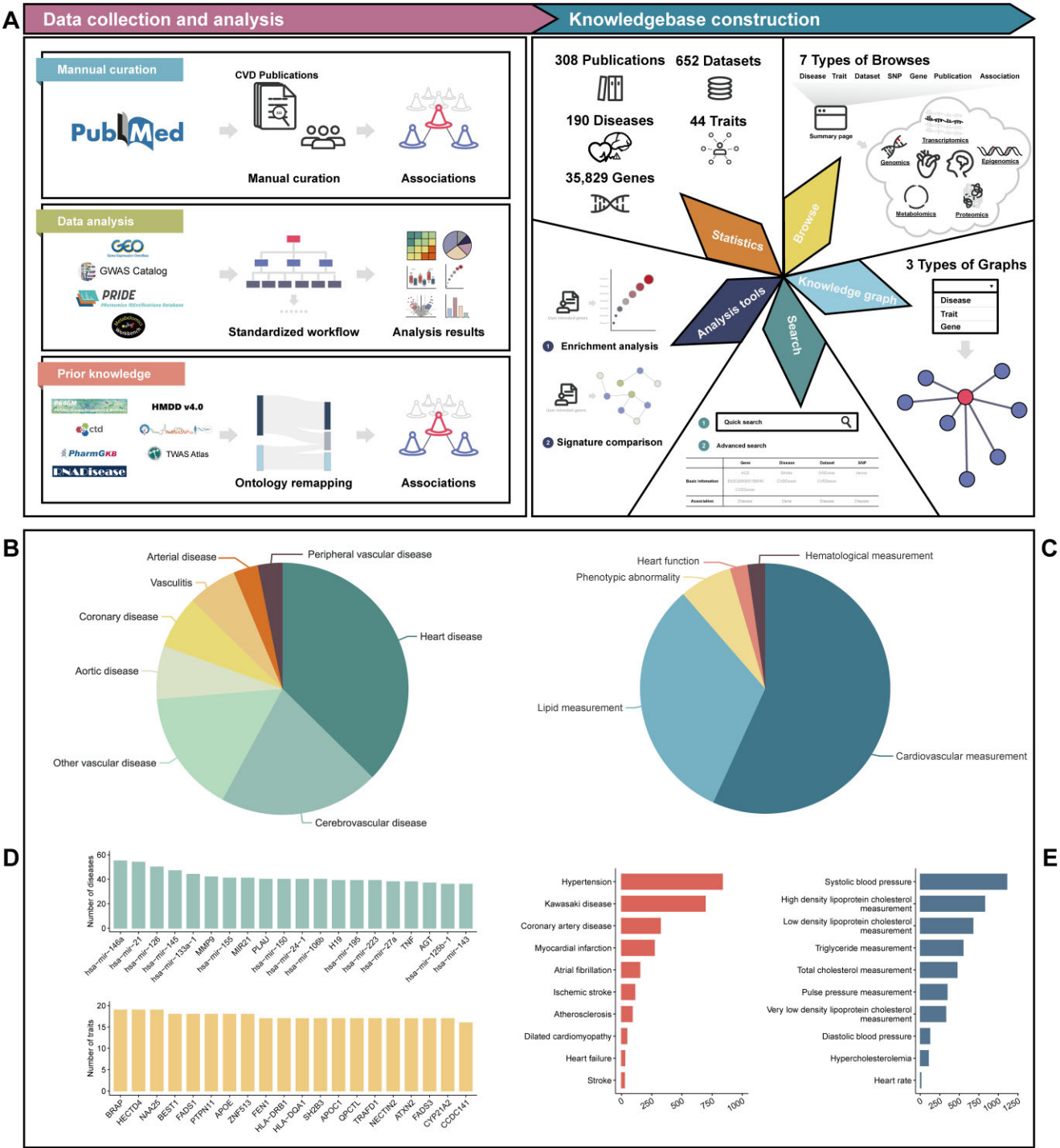


Figure 1. (A) An overview of CVD Atlas. **(B)** The distribution of 190 diseases among the 8 categories. **(C)** The distribution of 44 traits among the 5 categories. **(D)** The top 20 genes with the highest number of associated diseases (top) and traits (bottom). **(E)** The top 20 diseases (left) and traits (right) with the highest associated genes.

using specific terms related to genes (e.g. gene symbol), diseases or traits (e.g. disease or trait name), SNPs (e.g. dbSNP ID (49)) and datasets (e.g. GEO ID). The advanced search also enables users to search for genes, SNPs, or datasets associated with a specific disease or trait and to query diseases related to a specific gene (Figure 2B). Furthermore, CVD Atlas features an auto-suggestion function that provides candidate query terms for users based on even short inputs.

Highly integrated knowledge graph with interactive visualization

To improve the integration and visualization of information, CVD Atlas creates a comprehensive and interactive knowledge graph by systematically combining disease-gene and trait-gene associations (Figure 2C). The graph consists of three panels: ‘Disease network’, ‘Trait network’ and ‘Gene network’. It defines three types of entities: diseases, traits and

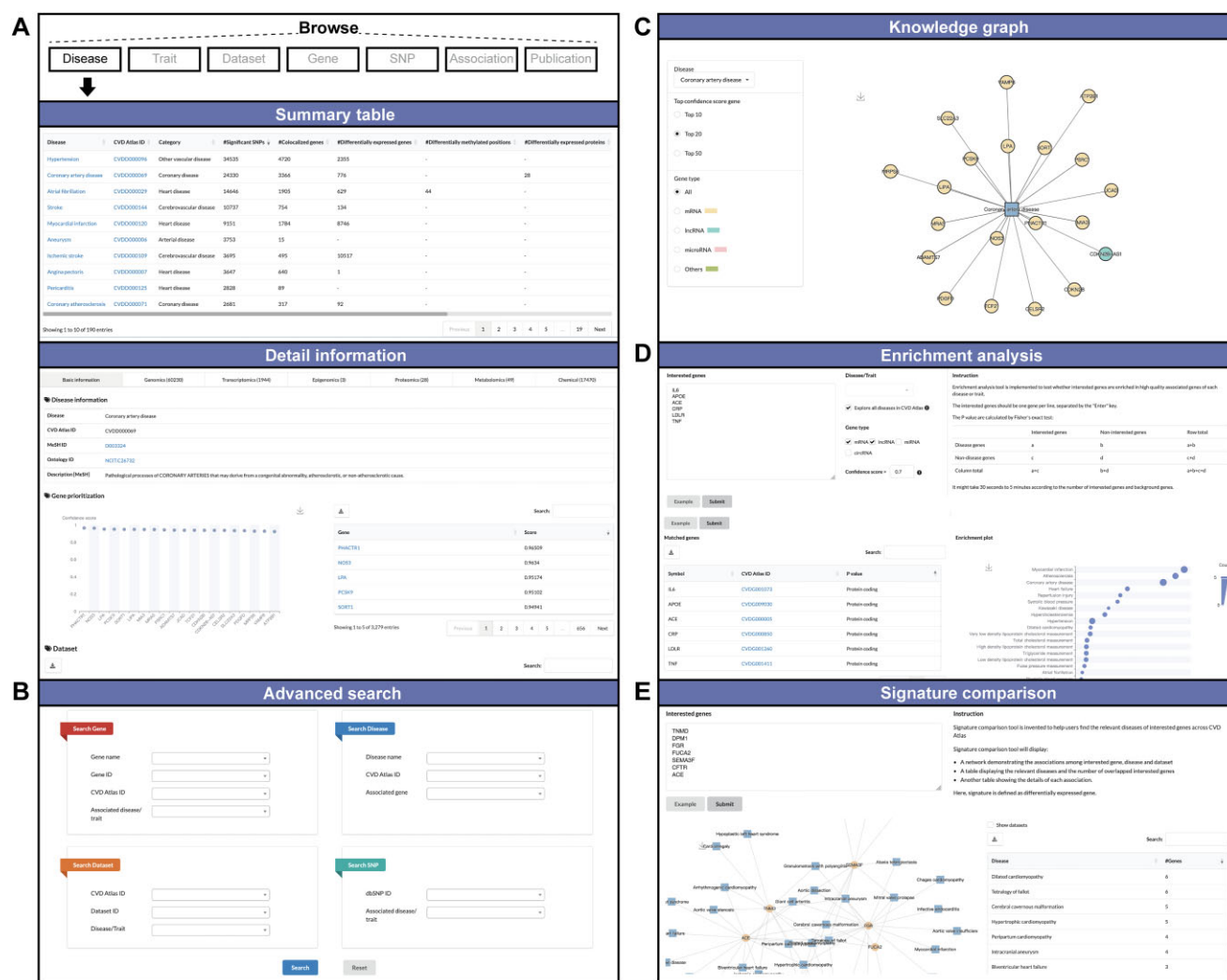


Figure 2. Illustration of module interfaces in CVD Atlas. (A) Screenshot of the browse table with fundamental information and summary statistics (top), using disease browse as an example, and detailed disease information, using coronary artery disease as an example, including basic information, genomics, transcriptomics, epigenomics, proteomics, metabolomics, chemical, and other information (bottom). (B) Screenshot of the advanced search module. (C) A screenshot of the knowledge graph using coronary artery disease as an example is included. (D, E) Screenshots of analysis tools for enrichment analysis (D) and signature comparison (E), respectively.

genes. Each panel highlights the respective entity as core nodes. The connections are quantitatively characterized by a confidence scoring system based on multi-omics associations and represented by the distance between two nodes. By default, only the top 20 associated entities are displayed and sorted in descending order by confidence score. Moreover, all nodes in the graph can be moved, allowing users to customize the display and export the high-resolution graph. For example, in the ‘Disease network’ for coronary artery disease, the graph displays the top 20 genes related to CAD by default, such as *PHACTR1* and *NOS3*. Notably, *PHACTR1* is the most related gene, as it is the nearest node to CAD, with a confidence score of 0.965. The graph lets users access detailed information about each node by clicking on them.

Analysis tools for efficient data exploration

We have implemented two tools, ‘Enrichment analysis’ and ‘Signature comparison’, in CVD Atlas to make it easy and efficient to explore the database. The first tool helps users determine if specific genes play a role in certain diseases or traits by examining which diseases or traits are enriched by the genes

of interest. Users can submit a list of genes under specified conditions to calculate the *P*-value using Fisher’s exact test. The ‘Enrichment analysis’ output includes a table for genes matched in CVD Atlas, a dot plot, and a detailed table of enrichment results (Figure 2D). The second tool helps users identify the relevant diseases for user-submitted gene signatures across datasets by assessing the overlap between the submitted gene list and the differentially expressed genes for diseases in each dataset. The ‘Signature comparison’ output consists of a network graph showing the relationships between genes, diseases, and datasets, along with two tables detailing the number of gene hits per disease and the differential analysis results (Figure 2E).

Case study: coronary artery disease

CVD Atlas contains comprehensive information about cardiovascular disease (CVD) and its functional modules. To illustrate, let’s look at how CVD Atlas is used to explore coronary artery disease (CAD). CAD is a cardiovascular disorder characterized by the narrowing or blockage of the coronary arteries. There are 79,724 entries related to CAD within CVD

Atlas across all six omics levels. CAD is linked with 3,279 genes, with 379 genes associated with CAD in at least two omics levels. The detailed information about CAD is divided into six panels: ‘Basic information’, ‘Genomics’, ‘Transcriptomics’, ‘Epigenomics’, ‘Proteomics’ and ‘Chemical’. The ‘Basic information’ panel provides fundamental details about CAD, such as the disease name, CVD Atlas ID, MeSH ID, ontology ID and a description. A confidence score presents an overview of the associations between genes and CAD. The top five genes associated with CAD are *PHACTR1*, *NOS3*, *LPA*, *PCSK9* and *SORT1* (Supplementary Figure S2). Taking *NOS3* as an example, it is linked to CAD across three omics levels: genomics, transcriptomics, and chemicals. This association is supported by 598 entries, covering 43 publications and two datasets. This strongly underscores that the integrated data provided by CVD Atlas offers users multidimensional insights into the connection between specific genes and diseases from different angles.

CVD Atlas provides six other panels to offer a comprehensive view. In the ‘Genomics’ panel, five GWAS datasets (four European and one mixed ancestry) were included. Among them, rs7412, a missense variant of *APOE*, has the highest number of significant datasets in the European population (Supplementary Figure S3), suggesting its potential importance for CAD. Colocalization analysis of GWAS and coronary artery eQTL data indicates that *PHACTR1*, *MRAS* and *LINC00881* might be strong causal candidates for CAD in the European population (Supplementary Figure S4). These genes are closely associated with CAD and other cardiovascular diseases (50–54). The ‘Transcriptomics’ panel displays differentially expressed genes collected from dataset analysis and manual curation. For example, *JUN*, *NR4A1* and *TNFAIP3* are significantly differentially expressed in three of ten datasets (Supplementary Figure S5). These genes have close relationships with human malignancies (55–59). Recent studies reported that CAD and cancer often co-occur via common biological pathways and shared risk factors. Various chemotherapeutic agents and radiotherapy can impact the development and progression of CAD (60). In the ‘Epigenomics’ panel, no differentially methylated positions are recorded in the current version of CVD Atlas. Moreover, associations at different omics levels collected from existing databases and curated results are provided in these three panels. In the ‘Proteomics’ panel, *IGFBP2*, *THY1* and *APOH* are the top positively related proteins, whereas *S100A8*, *TGM3*, and *APOB* are the top negatively associated proteins in CAD (Supplementary Table S1). In the ‘Metabolomics’ panel, 49 differentially expressed metabolites are displayed. Among them, adenylosuccinic acid, 4a-carbinolamine tetrahydrobiopterin, and succinyladenosine are the top metabolites associated with CAD (Supplementary Table S2). The associations of chemicals with diseases were presented in the ‘Chemical’ panel. For example, cocaine is the most closely related chemical, directly supported by 17 publications as the ‘Marker/Mechanism’ type in CTD. In comparison, aspirin is directly reported by eight publications as the ‘Therapeutic’ type (Supplementary Figure S6). CVD Atlas contains comprehensive information about cardiovascular diseases and offers valuable insights for researchers on cardiovascular health.

Discussion and future developments

Cardiovascular disease (CVD) is a significant health concern that has been extensively researched. Numerous studies

have been conducted on CVD across different fields, including genomics, transcriptomics, epigenomics, proteomics and metabolomics (61,62). Existing CVD multi-omics resources offer a comprehensive perspective and valuable insights into CVD. However, integrating a large number of associations and datasets remains a challenge. To address this issue, we introduce the CVD Atlas. CVD Atlas is an available database that systematically curates published disease-gene associations, analyzes multi-omics datasets, combines knowledge from existing databases, and implements a visualized and interactive knowledge graph and two analysis tools. Compared with the current CVD multi-omics databases, CVD Atlas offers the following features: (i) comprehensive coverage of a wide range of cardiovascular diseases grouped by an ontology classification system and characterized by multi-omics annotations; (ii) integration of analysis results from diverse omics datasets processed through a unified pipeline, providing valuable insights into the underlying mechanisms of diseases and (iii) aggregation of scattered knowledge from multiple existing databases to provide a more comprehensive understanding of CVD.

CVD Atlas, a crucial resource within the National Genomics Data Center (NGDC, <https://ngdc.cnecb.ac.cn>) (63) committed to continuous improvement, will be regularly updated to include the latest association discoveries and cardiovascular disease (CVD) datasets from various omics. The updates will expand the types of knowledge or datasets generated by different methods, including additional levels of epigenomic knowledge like histone modification and chromatin accessibility (64,65). Moreover, future updates will incorporate single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics data to analyze cellular transcriptomic variability under different conditions of CVD (66,67). The current version of CVD Atlas already includes one expression quantitative trait loci (eQTL) dataset from GTEx, and future updates will integrate more types of QTLs while considering the impact of ancestry. Additionally, regulatory information on protein–protein interaction, which is valuable for understanding the complex biological pathways and molecular mechanisms underlying CVD, will also be included in future updates. The aim is for CVD Atlas to become an essential resource, providing comprehensive and up-to-date knowledge on cardiovascular disease.

Data availability

CVD Atlas is available at <https://ngdc.cnecb.ac.cn/cvd/>.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank the users for reporting bugs and providing suggestions.

Funding

Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38030400 to J.X.]; National Natural Science Foundation of China [32170669 to J.X., 32300542 to J.Z.]; National Key Research Program of China [2020YFA0907001 to J.X.]; Youth Innovation Promotion Association of the Chinese Academy of Sciences [2022098 to J.Z.]. Funding for open

access charge: Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38030400 to J.X.]; National Natural Science Foundation of China [32170669 to J.X., 32300542 to J.Z.]; National Key Research Program of China [2020YFA0907001 to J.X.]; Youth Innovation Promotion Association of the Chinese Academy of Sciences [2022098 to J.Z.].

Conflict of interest statement

None declared.

References

- Collaborators,G.B.D.C.o.D. (2018) Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*, **392**, 1736–1788.
- Lindstrom,M., DeCleene,N., Dorsey,H., Fuster,V., Johnson,C.O., LeGrand,K.E., Mensah,G.A., Razo,C., Stark,B., Varieur Turco,J., *et al.* (2022) Global Burden of Cardiovascular Diseases and Risks Collaboration, 1990–2021. *J. Am. Coll. Cardiol.*, **80**, 2372–2425.
- Shiffman,D., Ellis,S.G., Rowland,C.M., Malloy,M.J., Luke,M.M., Jakoubova,O.A., Pullinger,C.R., Cassano,J., Aouizerat,B.E., Fenwick,R.G., *et al.* (2005) Identification of four gene variants associated with myocardial infarction. *Am. J. Hum. Genet.*, **77**, 596–605.
- Zhuo,L.A., Wen,Y.T., Wang,Y., Liang,Z.F., Wu,G., Nong,M.D. and Miao,L. (2019) LncRNA SNHG8 is identified as a key regulator of acute myocardial infarction by RNA-seq analysis. *Lipids Health Dis.*, **18**, 201.
- Zhu,Y., Wloch,A., Wu,Q., Peters,C., Pagenstecher,A., Bertalanffy,H. and Sure,U. (2009) Involvement of PTEN promoter methylation in cerebral cavernous malformations. *Stroke*, **40**, 820–826.
- Lygirov,V., Latosinska,A., Makridakis,M., Mullen,W., Delles,C., Schanstra,J.P., Zoidakis,J., Pieske,B., Mischak,H. and Vlahou,A. (2018) Plasma proteomic analysis reveals altered protein abundances in cardiovascular disease. *J. Transl. Med.*, **16**, 104.
- Ganna,A., Salihovic,S., Sundstrom,J., Broeckling,C.D., Hedman,A.K., Magnusson,P.K., Pedersen,N.L., Larsson,A., Siegbahn,A., Zilmer,M., *et al.* (2014) Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet.*, **10**, e1004801.
- Teslovich,T.M., Musunuru,K., Smith,A.V., Edmondson,A.C., Stylianou,I.M., Koseki,M., Pirruccello,J.P., Ripatti,S., Chasman,D.I., Willer,C.J., *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.
- Wan,E.Y.F., Fung,W.T., Schooling,C.M., Au Yeung,S.L., Kwok,M.K., Yu,E.Y.T., Wang,Y., Chan,E.W.Y., Wong,I.C.K. and Lam,C.L.K. (2021) Blood pressure and risk of cardiovascular disease in UK Biobank: a mendelian randomization study. *Hypertension*, **77**, 367–375.
- Vinayagamoorthy,N., Hu,H.J., Yim,S.H., Jung,S.H., Jo,J., Jee,S.H. and Chung,Y.J. (2014) New variants including ARG1 polymorphisms associated with C-reactive protein levels identified by genome-wide association and pathway analysis. *PLoS One*, **9**, e95866.
- Bober,M., Wiehe,K., Yung,C., Onal Suzek,T., Lin,M., Baumgartner,W. Jr. and Winslow,R. (2002) CaGE: cardiac gene expression knowledgebase. *Bioinformatics*, **18**, 1013–1014.
- Liu,H., Liu,W., Liao,Y., Cheng,L., Liu,Q., Ren,X., Shi,L., Tu,X., Wang,Q.K. and Guo,A.Y. (2011) CADgene: a comprehensive database for coronary artery disease genes. *Nucleic Acids Res.*, **39**, D991–D996.
- V.A., Nayar,P.G., Murugesan,R., Mary,B., P.D. and Ahmed,S.S. (2015) CardioGenBase: A literature based Multi-omics database for major cardiovascular diseases. *PLoS One*, **10**, e0143188.
- Fernandes,M., Patel,A. and Husi,H. (2018) C/VDdb: A multi-omics expression profiling database for a knowledge-driven approach in cardiovascular disease (CVD). *PLoS One*, **13**, e0207371.
- 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.
- Lu,M., Zhang,Y., Yang,F., Mai,J., Gao,Q., Xu,X., Kang,H., Hou,L., Shang,Y., Qain,Q., *et al.* (2023) TWAS Atlas: a curated knowledgebase of transcriptome-wide association studies. *Nucleic Acids Res.*, **51**, D1179–D1187.
- Cui,C., Zhong,B., Fan,R. and Cui,Q. (2024) HMDD v4.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.*, **52**, D1327–D1332.
- 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.
- Davis,A.P., Wieggers,T.C., Johnson,R.J., Sciaky,D., Wieggers,J. and Mattingly,C.J. (2023) Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Res.*, **51**, D1257–D1262.
- Whirl-Carrillo,M., Huddart,R., Gong,L., Sangkuhl,K., Thorn,C.F., Whaley,R. and Klein,T.E. (2021) An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.*, **110**, 563–572.
- Chen,J., Lin,J., Hu,Y., Ye,M., Yao,L., Wu,L., Zhang,W., Wang,M., Deng,T., Guo,F., *et al.* (2023) RNADisease v4.0: an updated resource of RNA-associated diseases, providing RNA-disease analysis, enrichment and prediction. *Nucleic Acids Res.*, **51**, D1397–D1404.
- Sollis,E., Mosaku,A., Abid,A., Buniello,A., Cerezo,M., Gil,L., Groza,T., Gunes,O., Hall,P., Hayhurst,J., *et al.* (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.*, **51**, D977–D985.
- Murphy,A.E., Schilder,B.M. and Skene,N.G. (2021) MungeSumstats: A bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics*, **37**, 4593–4596.
- Wallace,C. (2021) A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.*, **17**, e1009440.
- The GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
- Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
- 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.
- Davis,S. and Meltzer,P.S. (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, **23**, 1846–1847.
- Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
- Chen,S. (2023) Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta*, **2**, e107.
- Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
- Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinf.*, **12**, 323.

33. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
34. 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.
35. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.*, **9**, 559.
36. Tian, Y., Morris, T.J., Webster, A.P., Yang, Z., Beck, S., Feber, A. and Teschendorff, A.E. (2017) ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*, **33**, 3982–3984.
37. Morris, T.J., Butcher, L.M., Feber, A., Teschendorff, A.E., Chakravarthy, A.R., Wojdacz, T.K. and Beck, S. (2014) ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics*, **30**, 428–430.
38. Butcher, L.M. and Beck, S. (2015) Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data. *Methods*, **72**, 21–28.
39. Perez-Riverol, Y., Bai, J., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.*, **50**, D543–D552.
40. Zhang, X., Smits, A.H., van Tilburg, G.B., Ovaa, H., Huber, W. and Vermeulen, M. (2018) Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat. Protoc.*, **13**, 530–550.
41. Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., Edison, A., Fiehn, O., Higashi, R., Nair, K.S., *et al.* (2016) Metabolomics Workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.*, **44**, D463–D470.
42. Pang, Z., Xu, L., Viau, C., Lu, Y., Salavati, R., Basu, N. and Xia, J. (2024) MetaboAnalystR 4.0: a unified LC-MS workflow for global metabolomics. *Nat. Commun.*, **15**, 3675.
43. Lu, H., Cassis, L.A., Kooi, C.W. and Daugherty, A. (2016) Structure and functions of angiotensinogen. *Hypertens. Res.*, **39**, 492–500.
44. Jeunemaitre, X., Soubrier, F., Kotelevtsev, Y.V., Lifton, R.P., Williams, C.S., Charru, A., Hunt, S.C., Hopkins, P.N., Williams, R.R., Lalouel, J.M., *et al.* (1992) Molecular basis of human hypertension: role of angiotensinogen. *Cell*, **71**, 169–180.
45. Pan, W.H., Chen, J.W., Fann, C., Jou, Y.S. and Wu, S.Y. (2000) Linkage analysis with candidate genes: the Taiwan young-onset hypertension genetic study. *Hum. Genet.*, **107**, 210–215.
46. Sato, N., Katsuya, T., Nakagawa, T., Ishikawa, K., Fu, Y., Asai, T., Fukuda, M., Suzuki, F., Nakamura, Y., Higaki, J., *et al.* (2000) Nine polymorphisms of angiotensinogen gene in the susceptibility to essential hypertension. *Life Sci.*, **68**, 259–272.
47. Wang, L., Shen, C., Yang, S., Chen, Y., Guo, D., Jin, Y., He, L., Chen, J., Zhao, X., Zhao, H., *et al.* (2015) Association study of NOS3 gene polymorphisms and hypertension in the Han Chinese population. *Nitric Oxide*, **51**, 1–6.
48. Salvi, E., Kuznetsova, T., Thijs, L., Lupoli, S., Stolarz-Skrzypek, K., D'Avila, F., Tikhonoff, V., De Astis, S., Barcella, M., Seidlerova, J., *et al.* (2013) Target sequencing, cell experiments, and a population study establish endothelial nitric oxide synthase (eNOS) gene as hypertension susceptibility gene. *Hypertension*, **62**, 844–852.
49. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
50. Liao, X., Kennel, P.J., Liu, B., Nash, T.R., Zhuang, R.Z., Godier-Furnemont, A.F., Xue, C., Lu, R., Colombo, P.C., Uriel, N., *et al.* (2023) Effect of mechanical unloading on genome-wide DNA methylation profile of the failing human heart. *JCI Insight*, **8**, e161788.
51. Yari, A., Saleh-Gohari, N., Mirzaee, M., Hashemi, F. and Saedi, K. (2022) A study of associations between rs9349379 (PHACTR1), rs2891168 (CDKN2B-AS), rs11838776 (COL4A2) and rs4880 (SOD2) polymorphic variants and coronary artery disease in Iranian population. *Biochem. Genet.*, **60**, 106–126.
52. Chen, L., Qian, H., Luo, Z., Li, D., Xu, H., Chen, J., He, P., Zhou, X., Zhang, T., Chen, J., *et al.* (2019) PHACTR1 gene polymorphism with the risk of coronary artery disease in Chinese Han population. *Postgrad. Med. J.*, **95**, 67–71.
53. Schunkert, H., Konig, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C., *et al.* (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.*, **43**, 333–338.
54. Erdmann, J., Grosshennig, A., Braund, P.S., Konig, I.R., Hengstenberg, C., Hall, A.S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Tregouet, D.A., *et al.* (2009) New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.*, **41**, 280–282.
55. Luo, M., Wang, X., Wu, S., Yang, C., Su, Q., Huang, L., Fu, K., An, S., Xie, F., To, K.K.W., *et al.* (2023) A20 promotes colorectal cancer immune evasion by upregulating STC1 expression to block “eat-me” signal. *Signal Transduct. Target Ther.*, **8**, 312.
56. Zhu, P., Liu, G., Wang, X., Lu, J., Zhou, Y., Chen, S., Gao, Y., Wang, C., Yu, J., Sun, Y., *et al.* (2022) Transcription factor c-Jun modulates GLUT1 in glycolysis and breast cancer metastasis. *BMC Cancer*, **22**, 1283.
57. Mohankumar, K., Shrestha, R. and Safe, S. (2022) Nuclear receptor 4A1 (NR4A1) antagonists target paraspeckle component 1 (PSPC1) in cancer cells. *Mol. Carcinog.*, **61**, 73–84.
58. Wang, Y., Ren, X., Li, W., Cao, R., Liu, S., Jiang, L., Cheng, B. and Xia, J. (2021) SPDEF suppresses head and neck squamous cell carcinoma progression by transcriptionally activating NR4A1. *Int. J. Oral Sci.*, **13**, 33.
59. Breitenacker, K., Homolya, M., Luca, A.C., Lang, V., Trenk, C., Petroczi, G., Mohrher, J., Horvath, J., Moritsch, S., Haas, L., *et al.* (2021) Down-regulation of A20 promotes immune escape of lung adenocarcinomas. *Sci. Transl. Med.*, **13**, eabc3911.
60. Das, D., Asher, A. and Ghosh, A.K. (2019) Cancer and coronary artery disease: common associations, diagnosis and management challenges. *Curr. Treat. Options Oncol.*, **20**, 46.
61. Zhan, C., Tang, T., Wu, E., Zhang, Y., He, M., Wu, R., Bi, C., Wang, J., Zhang, Y. and Shen, B. (2023) From multi-omics approaches to personalized medicine in myocardial infarction. *Front. Cardiovasc. Med.*, **10**, 1250340.
62. Leon-Mimila, P., Wang, J. and Huertas-Vazquez, A. (2019) Relevance of multi-omics studies in cardiovascular diseases. *Front. Cardiovasc. Med.*, **6**, 91.
63. Members, C.-N. and PartnersPartners. (2024) Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024. *Nucleic Acids Res.*, **52**, D18–D32.
64. Wang, N., Wang, W., Wang, X., Mang, G., Chen, J., Yan, X., Tong, Z., Yang, Q., Wang, M., Chen, L., *et al.* (2022) Histone lactylation boosts reparative gene activation post-myocardial infarction. *Circ. Res.*, **131**, 893–908.
65. Turner, A.W., Hu, S.S., Mosquera, J.V., Ma, W.F., Hodonsky, C.J., Wong, D., Auguste, G., Song, Y., Sol-Church, K., Farber, E., *et al.* (2022) Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat. Genet.*, **54**, 804–816.
66. Wehrens, M., de Leeuw, A.E., Wright-Clark, M., Eding, J.E.C., Boogerd, C.J., Molenaar, B., van der Kraak, P.H., Kuster, D.W.D., van der Velden, J., Michels, M., *et al.* (2022) Single-cell transcriptomics provides insights into hypertrophic cardiomyopathy. *Cell Rep.*, **39**, 110809.
67. Zheng, K., Lin, L., Jiang, W., Chen, L., Zhang, X., Zhang, Q., Ren, Y. and Hao, J. (2022) Single-cell RNA-seq reveals the transcriptional landscape in ischemic stroke. *J. Cereb. Blood Flow Metab.*, **42**, 56–73.