Nucleic Acids Research, 2025, **53**, D1363–D1371
https://doi.org/10.1093/nar/gkae853
Advance access publication date: 1 October 2024
Database issue

DMRdb: a disease-centric Mendelian randomization database for systematically assessing causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases

Xiao Zheng, Zhihao Tian, Xiaohui Che, Xu Zhang, Yu Xiang, Zhijian Ge, Zhaoyu Zhai, Qinfeng Ma and Jianbo Pan*

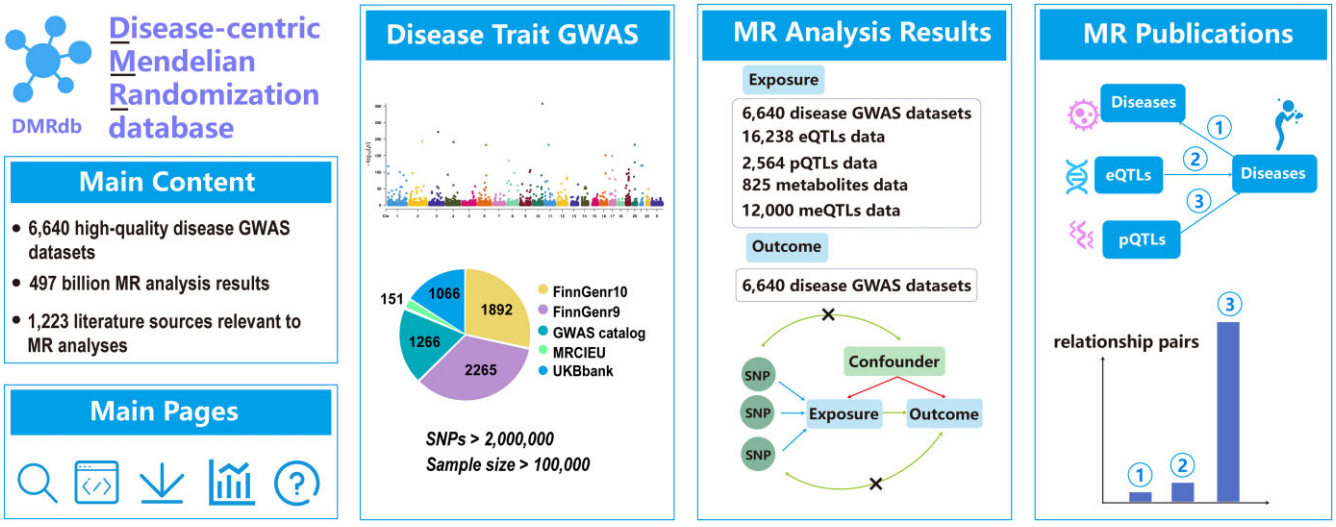
Basic Medicine Research and Innovation Center for Novel Target and Therapeutic Intervention, Ministry of Education, College of Pharmacy, and Precision Medicine Center, the Second Affiliated Hospital, Chongqing Medical University, Chongqing 400016, China

*To whom correspondence should be addressed. Tel: +86 23 684 80209; Fax: +86 23 684 80209; Email: panjianbo@cqmu.edu.cn

Abstract

Exploring the causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases is fundamental to the life sciences. However, large-scale research using Mendelian randomization (MR) analysis is currently lacking. To address this, we introduce DMRdb (http://www.inbirg.com/DMRdb/), a disease-centric Mendelian randomization database, designed to systematically assess causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases. The database consists of three main components: (i) 6640 high-quality disease genome-wide association studies (GWASs) from public sources that were subjected to rigorous quality filtering and standardization; (ii) over 497 billion results from MR analyses involving 6640 disease GWAS datasets, 16 238 expression quantitative trait loci (eQTLs) data, 2564 protein quantitative trait loci (pQTLs) data, 12 000 methylation quantitative trait locus (meQTLs) data and 825 metabolites data and (iii) over 380 000 causal relationship pairs from 1223 literature sources relevant to MR analyses. A user-friendly online database was developed to allow users to query, search, and download all the results. In summary, we anticipate that DMRdb will be a valuable resource for advancing our understanding of disease mechanisms and identifying new biomarkers and therapeutic targets.

Graphical abstract



Introduction

Diseases can be caused by various risk factors, including specific health conditions, genetic disorders, lifestyle choices and environmental factors. Understanding these causative factors can enhance our knowledge of disease entities, improve disease classification, and elucidate pathogenesis. In particular,

Received: August 6, 2024. **Revised:** September 10, 2024. **Editorial Decision:** September 17, 2024. **Accepted:** September 18, 2024

exploring the causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases can help identify biomarkers, therapeutic targets and gain in-depth knowledge of disease development mechanisms. Historically, observational epidemiology studies and randomized controlled trials (RCTs) have been used to make causal inferences. However, observational epidemiology has also produced high-profile failures where initially identified risk factors were later shown by RCTs to be noncausal. This discrepancy is often due to confounding factors, reverse causation, and selection bias (1–4), which are not fully addressed by statistical methods (5). Additionally, although RCTs are the gold standard for establishing causal relationships in health sciences, they have limitations, including ethical considerations, lengthy durations, limited generalizability and high costs and resource demands, making them feasible only with substantial preliminary evidence (6).

Mendelian randomization (MR) studies utilize genetic variations as proxies for modifiable environmental exposures to make causal inferences about the outcomes of these exposures. Genetic variants, such as single-nucleotide polymorphisms (SNPs), often obtained from genome-wide association studies (GWASs), expression quantitative trait loci (eQTLs), protein quantitative trait loci (pQTLs) and methylation quantitative trait loci (meQTLs) are linked to disease traits or affect gene or protein function and expression. These variants serve as instrumental variables (IVs) for MR, allowing the examination of the effects of perturbing specific targets. MR studies are akin to 'natural' RCTs (7,8), as they are based on Mendel's laws of inheritance and techniques for estimating IVs, facilitating the inference of causal effects even in the presence of unobserved confounders (9). MR can be used to provide insights into causality in situations where RCTs are not feasible, offering a reliable method for causal inference in observational studies (10). Owing to the numerous GWASs being conducted over the last decade, significant advancements in the methodology and application of MR have made it an essential tool for investigating the causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases to reveal risk factors, biomarkers and therapeutic targets for diseases. For example, researchers have utilized MR to determine the causal relationships of C-reactive protein (CRP) with metabolic syndrome traits (11,12), smoking and coronary stroke status (13), obesity and multiple sclerosis status (14), and depression and cardiovascular disease status (15). Despite its advantages, a central repository for comprehensive MR analysis results for all diseases is lacking, which hinders rapid access to information on disease associations.

To address these gaps, we developed a comprehensive resource called DMRdb (Disease-centric Mendelian Randomization database), available at http://www.inbirg.com/DMRdb/. DMRdb is used to curate the majority of published GWAS summary statistics for disease-related traits from multiple consortia through a series of filtering procedures, which are easy to download and can be applied for subsequent post-GWAS analyses by users. Additionally, DMRdb provides comprehensive MR analysis results, allowing users to query causal diseases, genes, proteins, CpG sites and metabolites. DMRdb also curates key conclusions from disease-related MR publications. In addition, DMRdb offers features such as rsID conversion and the analysis of phenomenon-wide association studies (PheWAS). In summary, we believe that DMRdb is a valuable resource of data on MR-based causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases for biomedical researchers.

Materials and methods

Data collection and curation

The DMRdb investigators continue to collect and curate GWAS summary-level association data for disease traits from online platforms such as the UK Biobank (https://www.nealelab.is/uk-biobank), MRC IEU (16), GWAS Catalog (17) and FinnGen (18). To standardize the format of GWAS datasets from these different platforms, we performed a series of quality control procedures. First, we checked the basic information of each dataset and removed redundant data between sources by using information on phenotypes, authors and publication sources. We then formatted the headers and ensured that the datasets contained the columns required for MR analysis, including SNPs, effect alleles, other alleles, effect allele frequency, beta, Standard Error (SE) and *P*-values. For data missing the required parameters, we calculated the missing statistics based on available statistics using the following approaches:

(1) Odds ratios (OR) Without Corresponding Beta Values: When the data includes OR value but misses corresponding beta value, the following formula is used to derive the beta:

$$\beta = \ln(\mathrm{OR})$$

(2) Calculating standard error (SE): If the data includes a 'beta' and '*P*-value' but lacks of 'SE', the standard error is calculated using the formula:

$$SE = \sqrt{\frac{\beta^2}{\chi^2_{1,upper}(p)}}$$

Here, 'SE' represents the standard error, 'β' is the beta estimate, and 'p' is the *P*-value. $\chi^2_{1,upper}(p)$ represents the upper tail quantile of the chi-squared distribution with 1 degree of freedom. Next, we supplemented incomplete datasets by adding missing SNP IDs, chromosome numbers, or base pair positions. We selected the reference genome (either hg19 or hg38) based on the version used in the original GWAS data. For datasets originally using GRCh37 (hg19), we employed the hg19 reference genome. Conversely, for data using hg38, we utilized the hg38 reference genome. The required reference genomes can be downloaded from the UCSC Genome Browser Downloads page (https://hgdownload.soe.ucsc.edu/downloads.html) or from the DMRdb Download page. We then filtered out data with fewer than 2 million SNPs and sample sizes <100 000. Finally, we saved all the data in a standardized format, as the downloaded GWAS summary statistics were saved in different formats depending on the platform or preprocessing process.

The summary data of the blood-based eQTL cohort of SNPs associated with the human transcriptome as a genetic tool come from the eQTLGen consortium (https://www.eqtlgen.org/), which comprises 37 datasets with a total of 31 684 blood samples, most of which were obtained from individuals of European ancestry (19). Summary data on SNPs associated with the human proteome as a genetic tool come from three large-scale GWASs. Sun *et al.* analysed 3622 plasma proteins in 3 301 individuals of European ancestry

(20); Folkersen *et al*. analysed 83 plasma proteins in 3394 European subjects (21); and Suhre *et al*. analysed 1 124 proteins in 1000 participants of European ancestry (22). For blood metabolites, Shin et al. reported on over 400 blood metabolites in 7824 adults from two European population studies (23); Kettunen et al. examined the genetic influences on 123 circulating metabolic traits in up to 24 925 individuals across ten European studies (24); and Borges *et al*. investigated 249 metabolic traits in a cohort of 115 078 participants using the Nightingale Health assay. Additionally, the summary data of CpG-associated meQTLs are sourced from the Genetics of DNA Methylation Consortium (GoDMC), a meQTL database containing genetic and methylation data from over 30 000 participants (25). The public data for the above QTLs and metabolites are accessible through the IEU GWAS database (https://gwas.mrcieu.ac.uk/) and GoDMC (http://mqtldb.godmc.org.uk/).

Data processing pipeline

MR analysis is a statistical method for inferring the causal relationship between an exposure and an outcome via GWAS summary data. For a genetic instrument to be considered valid, three basic assumptions must be met: (i) the genetic variants are associated with the exposure (relevance assumption); (ii) there are no unmeasured confounders affecting the associations between genetic variants and outcome (independence assumption) and (iii) the genetic variants affect the outcome only through their effect on the exposure (exclusion restriction) (9). When an MR study is performed, the selected IVs can serve as reliable proxies for the exposure by assumption 1, thereby addressing the constraints associated with directly utilizing the exposure variable. If the MR findings indicate a positive correlation and if there exists no direct link between IVs and the outcome as per assumption 2, the observed relationship can be attributed only to the exposure having a causal effect on the outcome, which constitutes the principal aim of MR analysis.

First, we analyzed GWAS summary data primarily from individuals of European ancestry to control for population differences. We constructed a rigorous framework to maintain the independence of SNPs associated with exposure, and we applied the criteria of $r^2 > 0.001$ and a clump window distance >10 000 kb. Additionally, a *P*-value < 5e-8 was set for the GWAS to ensure sufficient statistical power in the MR analysis. The effectiveness of the IVs was evaluated via the F statistic, with a minimum threshold of 10 for all the SNPs included in the study to minimize instrumental bias. SNPs related to the outcome (*P*-value < 1e-5) were excluded. Furthermore, we excluded palindromic SNPs to increase the robustness of the MR analysis. We subsequently used the R package TwoSampleMR to apply various well-established MR methods, including the Wald ratio, simple mode, simple median, MR–Egger, inverse-variance weighting (IVW; multiplicative random effects), IVW (fixed effects) and IVW. A scatter plot was generated to visualize the effect of exposure on the outcome. Next, we conducted a sensitivity analysis, including the generation of a leave-one-out plot and a forest plot, to evaluate the reliability of our findings. To address potential confounding factors in our MR analyses, we initially employed MR-Egger regression (26). This method plots the effect of SNPs on an exposure against their effect on an outcome. A non-zero intercept in this plot suggests the presence of pleiotropy, which we deemed significant if the *P*-value was <0.05. The slope provided by the MR-Egger regression offers adjusted causal estimates that are reliable even if some SNPs do not fully meet MR assumptions. Additionally, we utilized the weighted median approach (27) to enhance the precision of our estimates. In this approach, MR estimates are ordered and weighted according to the inverse of their variance. This method ensures that the median estimate remains robust, provided that over 50% of the weighting comes from SNPs that are not affected by pleiotropic effects. We also employed funnel plots as a visual method to check for symmetry among the SNPs. These plots are useful for identifying any asymmetry or outliers, which may indicate potential issues in our analysis. Together, these methods strengthen the validity of our findings by ensuring a comprehensive control of confounding variables and a robust analysis of the genetic data. Finally, we performed fine mapping analyses using the SuSiE method (28) to identify and prioritize the genetic variants most likely responsible for the observed disease traits.

Database construction

DMRdb is freely available at http://www.inbirg.com/DMRdb/. The online database framework was constructed via Django (https://docs.djangoproject.com/) and deployed on NGINX and uWSGI in a CentOS environment. All datasets are stored and managed with the PostgreSQL server (https://www.postgresql.org/) and filesystem. The frontend interface was developed via Vue3 (https://vuejs.org/). Frontend packages such as Element-plus (https://element-plus.org/) and ECharts (https://echarts.apache.org/) were used for the visual presentation of the results.

Database content and usage

Schematic overview

A schematic overview of DMRdb is illustrated in Figure 1. The current version of DMRdb includes summary data from 6640 disease GWASs totalling 4047 unique traits across 32 domains. Among these, 6246 (94%) GWAS studies involve SNP counts exceeding 10 billion and 5471 (82%) GWAS studies feature sample sizes exceeding 200 000 participants. These datasets collectively provide a valuable resource for researchers to explore causal variants and genes underlying various traits and conditions.

On the basis of the GWAS datasets mentioned above, we performed MR analyses to systematically identify potential causal relationships in five categories: disease−disease, gene−disease, protein−disease, metabolite-disease, and CpG site-disease. This analysis included 6640 disease GWAS datasets, 16 238 eQTLs data, 2564 pQTLs data, 825 metabolites and 12 000 meQTLs data. Our results included >497 million evaluated causal relationship pairs, with >38 million pairs showing a suggestive association (*P*-value < 0.05; 7.6%). The IVW method yielded 89 240 229 pairs, of which 4 482 339 (5.0%) had a significant *P*-value (<0.05). These results were visualized via relationship diagrams and scatter plots. In addition, sensitivity analyses, including Cochrane's *Q* test, MR–Egger's intercept test, and leave-one-out analysis, were performed, with results presented through forest plots, funnel plots, etc. Furthermore, conclusions from 1223 literature sources on MR analysis of diseases were carefully compiled, covering associations between diseases and

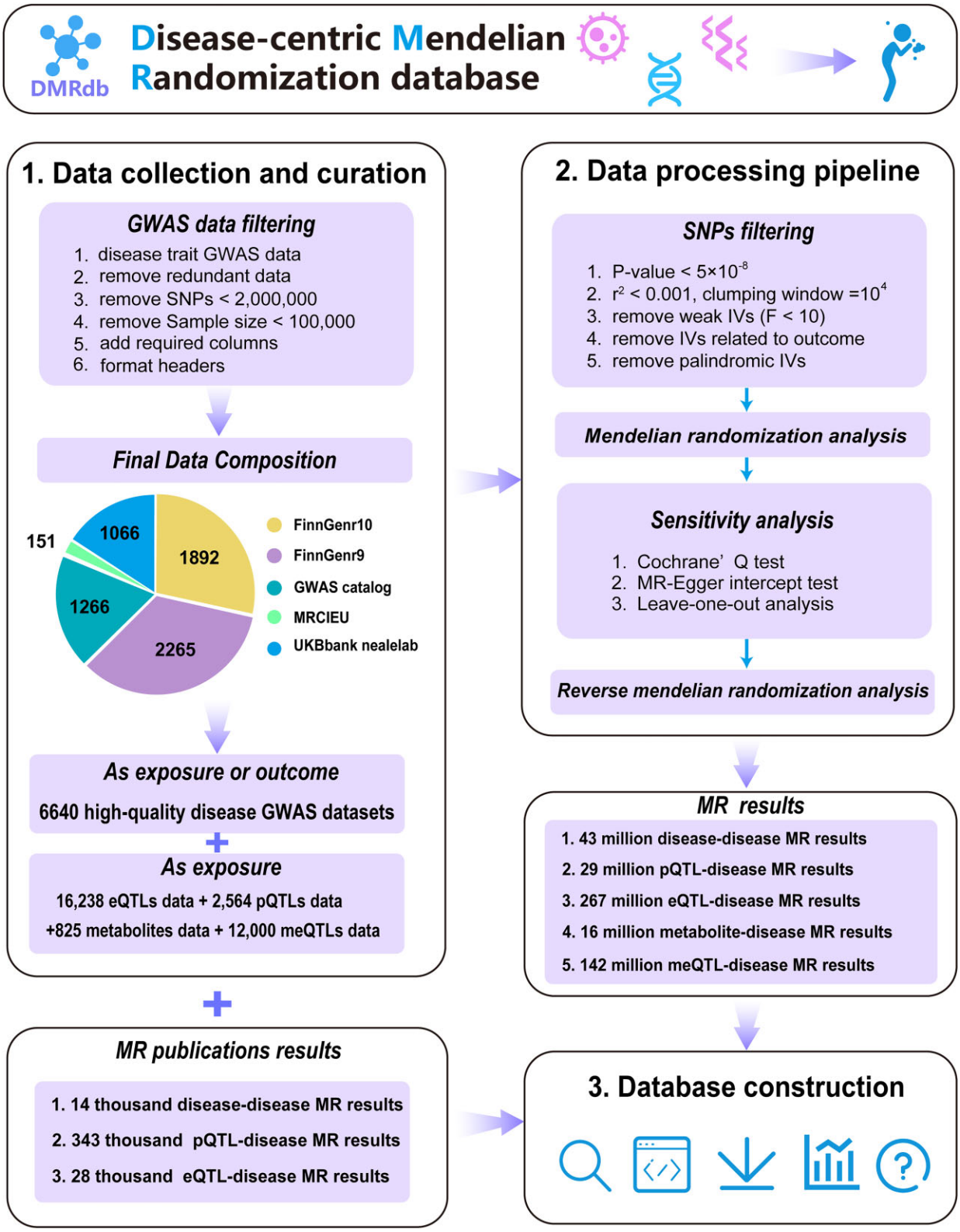


Figure 1. Schematic overview of DMRdb. This figure presents a detailed workflow and data processing pipeline for DMRdb. In the 'Data Collection and Curation' phase, 6640 high-quality disease GWAS datasets were filtered and standardized from sources including FinnGen, GWAS Catalog, MRC IEU and UK Biobank. During the 'Data Processing Pipeline' phase, SNPs underwent rigorous filtering, followed by MR and sensitivity analyses. This process generated over 497 million results, comprising 43 million disease–disease MR results, 29 million pQTL–disease MR results, 267 million eQTL–disease MR results, 142 million mQTL–disease MR results and 16 million metabolite–disease MR results. Additionally, over 380 thousand causal relationship pairs were collected from 1223 literature sources relevant to MR analyses. In the 'Database Construction' phase, a user-friendly online database was developed, enabling users to query, search, and download all the results.

eQTL/pQTL relationships between diseases, including 1057 publications on MR studies on causal relationships between diseases and 166 MR studies on eQTL/pQTL relationships between diseases, resulting in >380 thousand pairs of causal relationships.

Web interface

DMRdb provides user-friendly web interfaces that allow easy access to a wealth of information about disease-related GWAS data and MR analysis results. These interfaces include a range of functions for searching, browsing and downloading the data.

On the Home page, users can enter keywords such as disease names (e.g. COVID-19), gene names (e.g. CA4) or protein names (e.g. galectin-9) via the quick search function (Figure 2A). The search results were derived via the IVW method. For users looking for more detailed information, an 'Advanced Search' option is available on the 'Search' page (Figure 2B). This option allows one to filter disease records by various characteristics, including 'Result type', 'Analysis type' and 'Methods'. The input fields for the exposure and outcome names contain a tree selector that searches for the best match on the basis of the user input. Users can select multiple exposures or outcomes simultaneously. By applying certain criteria, disease records that match the selected characteristics are displayed on the results page.

The Browse page consists of a sidebar and a data table (Figure 2C). On the sidebar, users can access the MR analysis results in an interactive table. Users can customize filters to search for MR analysis results on the basis of different exposures, such as diseases, genes, proteins, CpG sites and metabolites, and can select multiple methods. The data table presents the results of the MR analysis. Clicking on 'ID' allows the user to view the corresponding detailed information page, and clicking on 'Plot' allows the user to view a diagram of the causal relationship on the basis of the page's content. When users click on the name of a method in the diagram, it hides the relationship line associated with that method. The 'Search' button can be clicked to display all custom relationship pairs on the canvas, and the download icon can be clicked to download the page's content. For more comprehensive results, users can download all the results from the download section on the Download page.

When a user clicks on a study ID on the Results or Browse page, they are redirected to a detailed page of the corresponding dataset. This detailed page provides comprehensive information and visualized charts related to the MR analysis (Figure 2D). The 'Scatter plot' option shows the associations between SNPs and outcomes plotted against the associations between SNPs and exposure, providing an immediate picture of the causal effect estimate for each variant. The 'Leave-one-out analysis' option shows the effect calculated via the IVW method after each SNP is removed, observing whether the beta values are all >0 or <0. Consistent directions indicate that a positive causal relationship between exposure and outcome remains even after removing certain SNPs. The 'MR−Egger and IVW test' results indicate the presence or absence of heterogeneity, with a *P*-value >0.05 suggesting that there was no heterogeneity. The results of the 'Pleiotropy test' indicate the presence or absence of horizontal pleiotropy, with a *P*-value >0.05 indicating that there was no horizontal pleiotropy. The 'Single SNP plot' function tests whether the conclusion of the study is strongly influenced by a single SNP by removing the SNPs one by one and observing changes in the effect value, reflecting the robustness of the conclusion. The funnel plot focuses on the symmetry of points on the left and right sides of the IVW line. Outlier points, such as those on the far side, indicate the presence of outliers. The 'Radial plot' function can be used to identify outlier variants on the basis of their contribution to global heterogeneity. The 'Harmonized data of MR' function shows the raw data used to perform the MR analysis, whereas 'The results of MR analysis' function shows all the results obtained via multiple methods. All the results can be downloaded by clicking on the download icon. The Disease GWAS page shows the basic information about disease trait GWAS datasets (Figure 2E). When a user clicks on a GWAS ID, they are redirected to a detailed page of the corresponding dataset. This detailed page provides comprehensive information about the GWAS data, including the individual SNPs (*P*-value < 5e-8), QQ plot, Manhattan plot, etc. (Figure 2F).

The tools page contains various instruments that are useful for the MR analysis workflow. Users can calculate F statistics by entering or uploading CSV files. Additionally, users can convert chr:pos (chromosome and position) to rsID (reference SNP ID) or vice versa. Moreover, the power value can be calculated on the basis of the sample size or, given a specific power value, the sample size required to achieve that value can be calculated. Users can utilize Disease PheWAS analysis by entering any SNP of interest and querying its *P*-value across all disease records. If the SNP is empirically associated with a phenotype, then this function will establish a link between that phenotype and all disease phenotypes. The scatter plot displays these results, with values >100 capped at 100. The links at the bottom of the plot refer to external websites with more detailed information about the SNP.

Application examples of DMRdb

In observational studies, type 2 diabetes (T2D) status is associated with a greater risk of developing coronary heart disease (CHD) (29,30). However, observational epidemiological studies of this nature are prone to various biases. Distinguishing causal relationships from those influenced by confounding factors or reverse causation is challenging. In this study, we used MR to obtain unconfounded estimates of the influence of genetically predicted T2D status on CHD risk. The MR approach involves the use of genetic variants as IVs to infer causality, thereby minimizing the biases associated with traditional observational studies (31).

The specific steps are as follows:

(1) Dataset selection:
   - × Exposure: Disease GWAS ID: DGWAS-5576 (Trait: T2D; Sample size: 655 666).
   - × Outcome: Disease GWAS ID: DGWAS-5735 (Trait: CHD; Sample size: 194 427).

(2) Analysis execution:
   - × Enter the Exposure ID and Outcome ID on the Search page.
   - × Select all available MR methods.
   - × Click the 'Submit' button to query the results.

The MR analysis results indicated that genetically predicted T2D status increased the risk of developing CHD. This is evidenced by an odds ratio (OR) of 1.10 (95% confidence

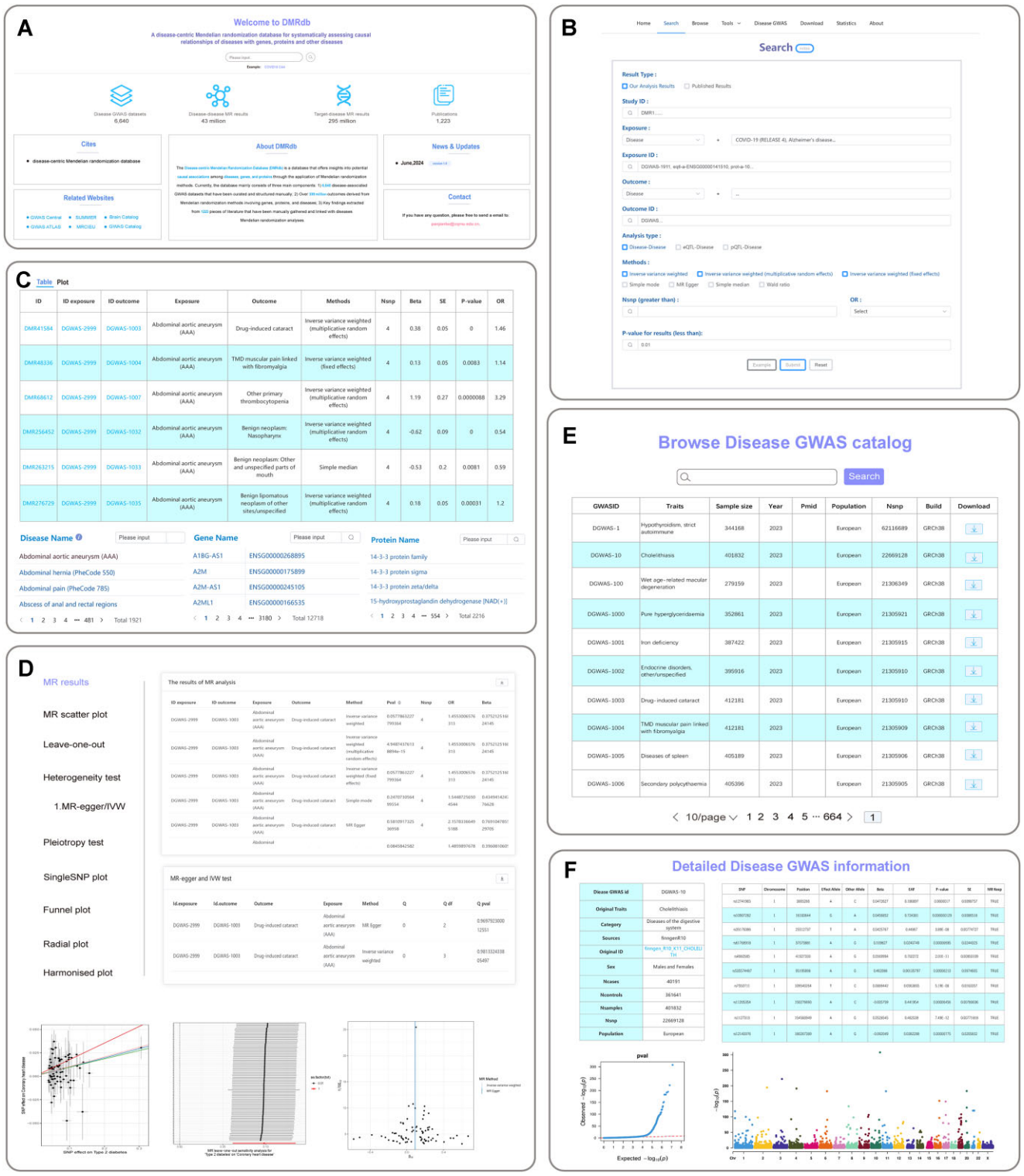


Figure 2. Screenshots of DMRdb web pages. (**A**) The search function on the homepage allows users to search quickly for MR results across multiple modules, including diseases, genes, and proteins. (**B**) The Search page provides an advanced search function to further refine the MR results. (**C**) Browse page mainly consists of sidebar and data table. The list of MR analysis results could be viewed in an interactive table on this page. On the sidebar, users can customize filters to search for MR analysis results on the basis of various exposures, such as diseases, genes and proteins. (**D**) Detailed information on the MR results is displayed when users click the Study ID. (**E**) On the Disease GWAS page, users can browse basic information about disease trait GWAS datasets. They can download GWAS data by clicking the download button. (**F**) Detailed information about a disease GWAS dataset is displayed when users click the Study ID.

ID.exposure	Exposure	ID.outcome	Outcome	Methods	Nsnp	P−value	OR (95% CI)
DGWAS−5576	Type 2 diabetes	DGWAS−5735	Coronary heart disease	Inverse variance weighted	75	1.41e−07	1.10 (1.06 − 1.14)
				Inverse variance weighted (multiplicative random effects)	75	1.41e−07	1.10 (1.06 − 1.14)
				Inverse variance weighted (fixed effects)	75	1.00e−07	1.10 (1.06 − 1.14)
				Simple mode	75	2.98e−03	1.19 (1.07 − 1.34)
				MR Egger	75	2.96e−02	1.09 (1.01 − 1.18)
				Simple median	75	5.85e−04	1.11 (1.05 − 1.18)
DGWAS−5735	Coronary heart disease	DGWAS−5576	Type 2 diabetes	Inverse variance weighted	8	1.90e−01	0.93 (0.82 − 1.04)
				Inverse variance weighted (multiplicative random effects)	8	1.90e−01	0.93 (0.82 − 1.04)
				Inverse variance weighted (fixed effects)	8	2.00e−02	0.93 (0.87 − 0.99)
				Simple mode	8	1.90e−01	0.85 (0.68 − 1.06)
				MR Egger	8	1.40e−01	0.60 (0.33 − 1.09)
				Simple median	8	1.00e−02	0.88 (0.80 − 0.97)

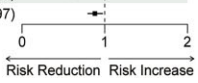


Figure 3. Potential causal associations of type 2 diabetes and coronary heart disease in two-sample bidirectional Mendelian randomization (MR) analyses.

interval [CI]: 1.06–1.14, *P*-value = 1.41e-7). Other MR methods corroborated this conclusion (Figure 3). By clicking on the study ID, we accessed detailed information on the MR analysis. This included additional analyses such as sensitivity analysis, heterogeneity testing, and horizontal pleiotropy testing to assess potential biases in the MR results. The results indicated that there was no significant bias (Supplementary Table S1). Additionally, reverse MR analysis was conducted to examine the influence of genetically predicted CHD status on the risk of developing T2D. However, no statistically significant associations were observed (Figure 3).

Further validation of our findings was performed, which reinforced the conclusion that T2D negatively affects CHD (Supplementary Figure S1). Our results are consistent with the findings of a publication titled 'A Mendelian randomization study of the effect of type 2 diabetes on coronary heart disease' (32). In that study, researchers reported that genetically predicted T2D status increases CHD risk, with an OR of 1.11 (95% CI: 1.05–1.17, *P*-value = 8.8e-5), on the basis of the IVW (multiplicative random effects) method. This result was added to our database and can be found by searching for 'Study ID: DMR363300926'.

In conclusion, these findings suggest that T2D has a negative effect on the occurrence of CHD. The consistent results across multiple MR methods and additional validation steps strengthen the evidence of a causal relationship between T2D status and increased CHD risk.

Discussion and future directions

Exploring the causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases can help identify biomarkers and therapeutic targets and provide in-depth knowledge of disease development mechanisms. In recent years, MR has emerged as a robust method that uses GWAS summary data to understand genetic influences on outcomes. Because genetic variants are inherited randomly in the germline, MR mitigates confounding factors that influence the associations between exposure and outcome in epidemiologic approaches (33).

Currently, databases such as the GWAS ATLAS (34) and GWAS Catalog (17) serve as GWAS data repositories. However, poor data quality and inconsistent data formats impede the direct use of these data on a large scale. Presently, a single database for aggregating standardized disease-related GWAS data comprehensively for large-scale MR analysis is lacking. Additionally, databases such as SUMMER (35) and Brain Catalogue (36) provide MR analysis results but focus on certain types of diseases and lack causal relationship investigations between diseases. SUMMER is dedicated to investigating 17 distinct cancer types with 150 risk factors and circulating biomarkers, whereas Brain Catalogue focuses on 517 brain disease traits with 58 QTL datasets. Platforms such as MR-Base (37) allow MR analyses but require significant prior knowledge and time, which is often hindered by unstable server performance. Currently, there is a need for an enormous workload in data collection and high computational cost, as no platform comprehensively offers rapid access and visualization of pan-disease cross-population MR analysis results on the basis of various disease GWAS datasets.

To address these gaps, we developed DMRdb, a curated resource that provides high-quality GWAS datasets related to various types of diseases and MR results to explore causal relationships between diseases, genes, proteins, CpG sites and metabolites. Compared with existing MR-related databases such as SUMMER and Brain Catalog, DMRdb stands out for its comprehensive disease GWAS datasets and extensive MR analysis results, facilitating intuitive queries and straightforward data downloads. As the largest and most comprehensive database of its kind, DMRdb enables exploration of disease mechanisms and identification of therapeutic targets through MR analysis, etc.

Further enhancements to DMRdb include the expansion of disease GWAS datasets from multiple sources and the use of MR analysis to investigate more potential causal relationships between various exposures (e.g. the gut microbiome and immune cells) and diseases. DMRdb currently focuses predominantly on eQTL, pQTL and meQTL data from blood-related sources. Moving forward, we plan to diversify the database to include diverse tissue and omics datasets to increase its relevance and broaden its applicability in genetic research. Furthermore, we will develop a tool to allow users to upload their own molecular quantitative trait loci (xQTL) data for online MR analysis to enhance the functionality and collaborative potential of DMRdb. Additionally, we intend to implement more advanced analytical approaches, such as summary data-based MR analysis (SMR) (38), Bayesian colocalization analysis (COLOC) (39), transcriptome-wide association analysis (TWAS) (40), and stratified LD score regression (S-LDSC) (41), to gain deeper insights. The aim of these efforts is to refine the understanding of genetic influences on diseases and increase the utility of resources for biomedical

research by supporting investigations into relevant causal relationships, genetic mechanisms, biomarkers and therapeutic strategies.

In summary, DMRdb is an important resource for biomedical researchers seeking MR-based insights into disease causation, genetic interactions and therapeutic targets. It supports a wide range of research activities, including meta-analyses; association studies between diseases; elucidation of genetic mechanisms; identification of risk factors, biomarkers and therapeutic targets; and guidance for functional experiments.

Data availability

DMRdb is freely available at http://www.inbirg.com/DMRdb/.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The computing work in this paper was partly supported by Supercomputing Center of Chongqing Medical University.

Funding

National Natural Science Foundation of China [32470699]; Natural Science Foundation of Chongqing, China [CSTB2023NSCQ-MSX0289]; Research Startup Funds of Chongqing Medical University; University Innovation Research Group Project of Chongqing [CXQT21016]. Funding for open access charge: Chongqing Medical University.

Conflict of interest statement

None declared.

References

1. Lawlor,D.A., Davey Smith,G., Kundu,D., Bruckdorfer,K.R. and Ebrahim,S. (2004) Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet*, **363**, 1724–1727.
2. Vandenbroucke,J.P. (2004) Commentary: the HRT story: vindication of old epidemiological theory. *Int. J. Epidemiol.*, **33**, 456–457.
3. Vandenbroucke,J.P. (2004) When are observational studies as credible as randomised trials? *Lancet*, **363**, 1728–1731.
4. Lawlor,D.A. and Smith,G.D. (2006) Cardiovascular risk and hormone replacement therapy. *Curr. Opin. Obstet. Gynecol.*, **18**, 658–665.
5. Phillips,A.N. and Smith,G.D. (1992) Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat. Med.*, **11**, 953–961.
6. Lawlor,D.A., Harbord,R.M., Sterne,J.A., Timpson,N. and Davey Smith,G. (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.*, **27**, 1133–1163.
7. Hingorani,A. and Humphries,S. (2005) Nature's randomised trials. *Lancet*, **366**, 1906–1908.
8. Davey Smith,G. and Ebrahim,S. (2005) What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ*, **330**, 1076–1079.
9. Sanderson,E., Glymour,M.M., Holmes,M.V., Kang,H., Morrison,J., Munafò,M.R., Palmer,T., Schooling,C.M., Wallace,C., Zhao,Q., *et al.* (2022) Mendelian randomization. *Nat. Rev. Methods Primers*, **2**, 6.
10. Richmond,R.C. and Davey Smith,G. (2022) Mendelian randomization: concepts and scope. *Cold Spring Harb. Perspect. Med.*, **12**, a040501.
11. Davey Smith,G., Lawlor,D.A., Harbord,R., Timpson,N., Rumley,A., Lowe,G.D., Day,I.N. and Ebrahim,S. (2005) Association of C-reactive protein with blood pressure and hypertension: life course confounding and mendelian randomization tests of causality. *Arterioscler. Thromb. Vasc. Biol.*, **25**, 1051–1056.
12. Timpson,N.J., Lawlor,D.A., Harbord,R.M., Gaunt,T.R., Day,I.N., Palmer,L.J., Hattersley,A.T., Ebrahim,S., Lowe,G.D., Rumley,A., *et al.* (2005) C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *Lancet*, **366**, 1954–1959.
13. Larsson,S.C., Burgess,S. and Michaëlsson,K. (2019) Smoking and stroke: a mendelian randomization study. *Ann. Neurol.*, **86**, 468–471.
14. Mokry,L.E., Ross,S., Timpson,N.J., Sawcer,S., Davey Smith,G. and Richards,J.B. (2016) Obesity and multiple sclerosis: a Mendelian randomization study. *PLoS Med.*, **13**, e1002053.
15. Li,G.H., Cheung,C.L., Chung,A.K., Cheung,B.M., Wong,I.C., Fok,M.L.Y., Au,P.C. and Sham,P.C. (2022) Evaluation of bi-directional causal association between depression and cardiovascular diseases: a Mendelian randomization study. *Psychol. Med.*, **52**, 1765–1776.
16. Elsworth,B., Lyon,M., Alexander,T., Liu,Y., Matthews,P., Hallett,J., Bates,P., Palmer,T., Haberland,V., Smith,G.D., *et al.* (2020) The MRC IEU OpenGWAS data infrastructure. bioRxiv doi: https://doi.org/10.1101/2020.08.10.244293, 10 August 2020, preprint: not peer reviewed.
17. Sollis,E., Mosaku,A., Abid,A., Buniello,A., Cerezo,M., Gil,L., Groza,T., Güneş,O., Hall,P., Hayhurst,J., *et al.* (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.*, **51**, D977–D985.
18. Kurki,M.I., Karjalainen,J., Palta,P., Sipilä,T.P., Kristiansson,K., Donner,K.M., Reeve,M.P., Laivuori,H., Aavikko,M., Kaunisto,M.A., *et al.* (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*, **613**, 508–518.
19. Võsa,U., Claringbould,A., Westra,H.J., Bonder,M.J., Deelen,P., Zeng,B., Kirsten,H., Saha,A., Kreuzhuber,R., Yazar,S., *et al.* (2021) Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.*, **53**, 1300–1310.
20. Sun,B.B., Maranville,J.C., Peters,J.E., Stacey,D., Staley,J.R., Blackshaw,J., Burgess,S., Jiang,T., Paige,E., Surendran,P., *et al.* (2018) Genomic atlas of the human plasma proteome. *Nature*, **558**, 73–79.
21. Folkersen,L., Fauman,E., Sabater-Lleal,M., Strawbridge,R.J., Frånberg,M., Sennblad,B., Baldassarre,D., Veglia,F., Humphries,S.E., Rauramaa,R., *et al.* (2017) Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLos Genet.*, **13**, e1006706.
22. Suhre,K., Arnold,M., Bhagwat,A.M., Cotton,R.J., Engelke,R., Raffler,J., Sarwath,H., Thareja,G., Wahl,A., DeLisle,R.K., *et al.* (2017) Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.*, **8**, 14357.
23. Shin,S.Y., Fauman,E.B., Petersen,A.K., Krumsiek,J., Santos,R., Huang,J., Arnold,M., Erte,I., Forgetta,V., Yang,T.P., *et al.* (2014) An atlas of genetic influences on human blood metabolites. *Nat. Genet.*, **46**, 543–550.
24. Kettunen,J., Demirkan,A., Würtz,P., Draisma,H.H., Haller,T., Rawal,R., Vaarhorst,A., Kangas,A.J., Lyytikäinen,L.P., Pirinen,M., *et al.* (2016) Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.*, **7**, 11122.

25. 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.
26. Bowden,J., Davey Smith,G. and Burgess,S. (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.*, **44**, 512–525.
27. Bowden,J., Davey Smith,G., Haycock,P.C. and Burgess,S. (2016) Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.*, **40**, 304–314.
28. Wang,G., Sarkar,A., Carbonetto,P. and Stephens,M. (2020) A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.*, **82**, 1273–1300.
29. Sarwar,N., Gao,P., Seshasai,S.R., Gobin,R., Kaptoge,S., Di Angelantonio,E., Ingelsson,E., Lawlor,D.A., Selvin,E., Stampfer,M., *et al.* (2010) Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet*, **375**, 2215–2222.
30. Bhattacharyya,O.K., Shah,B.R. and Booth,G.L. (2008) Management of cardiovascular disease in patients with diabetes: the 2008 Canadian Diabetes Association guidelines. *CMAJ*, **179**, 920–926.
31. Smith,G.D. and Ebrahim,S. (2004) Mendelian randomization: prospects, potentials, and limitations. *Int. J. Epidemiol.*, **33**, 30–42.
32. Ahmad,O.S., Morris,J.A., Mujammami,M., Forgetta,V., Leong,A., Li,R., Turgeon,M., Greenwood,C.M., Thanassoulis,G., Meigs,J.B., *et al.* (2015) A Mendelian randomization study of the effect of type-2 diabetes on coronary heart disease. *Nat. Commun.*, **6**, 7060.
33. Sekula,P., Del Greco,M.F., Pattaro,C. and Köttgen,A. (2016) Mendelian randomization as an approach to assess causality using observational data. *J. Am. Soc. Nephrol.*, **27**, 3253–3265.
34. Watanabe,K., Stringer,S., Frei,O., Umićević Mirkov,M., de Leeuw,C., Polderman,T.J.C., van der Sluis,S., Andreassen,O.A., Neale,B.M. and Posthuma,D. (2019) A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.*, **51**, 1339–1348.
35. Xin,J., Gu,D., Chen,S., Ben,S., Li,H., Zhang,Z., Du,M. and Wang,M. (2023) SUMMER: a Mendelian randomization interactive server to systematically evaluate the causal effects of risk factors and circulating biomarkers on pan-cancer survival. *Nucleic Acids Res.*, **51**, D1160–D1167.
36. Pan,S., Kang,H., Liu,X., Lin,S., Yuan,N., Zhang,Z., Bao,Y. and Jia,P. (2023) Brain Catalog: a comprehensive resource for the genetic landscape of brain-related traits. *Nucleic Acids Res.*, **51**, D835–D844.
37. Hemani,G., Zheng,J., Elsworth,B., Wade,K.H., Haberland,V., Baird,D., Laurin,C., Burgess,S., Bowden,J., Langdon,R., *et al.* (2018) The MR-Base platform supports systematic causal inference across the human phenome. *eLife*, **7**, e34408.
38. Zhu,Z., Zhang,F., Hu,H., Bakshi,A., Robinson,M.R., Powell,J.E., Montgomery,G.W., Goddard,M.E., Wray,N.R., Visscher,P.M., *et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.*, **48**, 481–487.
39. Giambartolomei,C., Vukcevic,D., Schadt,E.E., Franke,L., Hingorani,A.D., Wallace,C. and Plagnol,V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.
40. Barbeira,A.N., Dickinson,S.P., Bonazzola,R., Zheng,J., Wheeler,H.E., Torres,J.M., Torstenson,E.S., Shah,K.P., Garcia,T., Edwards,T.L., *et al.* (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.*, **9**, 1825.
41. Finucane,H.K., Bulik-Sullivan,B., Gusev,A., Trynka,G., Reshef,Y., Loh,P.R., Anttila,V., Xu,H., Zang,C., Farh,K., *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.*, **47**, 1228–1235.

Received: August 6, 2024. **Revised:** September 10, 2024. **Editorial Decision:** September 17, 2024. **Accepted:** September 18, 2024