Published online 16 October 2022 *Nucleic Acids Research, 2023, Vol. 51, Database issue* ***D835–D844***
https://doi.org/10.1093/nar/gkac895

Brain Catalog: a comprehensive resource for the genetic landscape of brain-related traits

Siyu Pan[1,2,†], Hongen Kang[1,2,†], Xinxuan Liu[1,3,†], Shiqi Lin[1,2], Na Yuan[1], Zhang Zhang[4,*], Yiming Bao[3,4,*] and Peilin Jia[1,*]

[1]CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China, [2]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100101, China, [3]School of Future Technology, University of Chinese Academy of Sciences, Beijing 100101, China and [4]National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China

Received August 15, 2022; Revised September 13, 2022; Editorial Decision September 25, 2022; Accepted October 08, 2022

ABSTRACT

A broad range of complex phenotypes are related to dysfunctions in brain (hereafter referred to as brain-related traits), including various mental and behavioral disorders and diseases of the nervous system. These traits in general share overlapping symptoms, pathogenesis, and genetic components. Here, we present Brain Catalog (https://ngdc.cncb.ac.cn/braincatalog), a comprehensive database aiming to delineate the genetic components of more than 500 GWAS summary statistics datasets for brain-related traits from multiple aspects. First, Brain Catalog provides results of candidate causal variants, causal genes, and functional tissues and cell types for each trait identified by multiple methods using comprehensive annotation datasets (58 QTL datasets spanning 6 types of QTLs). Second, Brain Catalog estimates the SNP-based heritability, the partitioning heritability based on functional annotations, and genetic correlations among traits. Finally, through bidirectional Mendelian randomization analyses, Brain Catalog presents inference of risk factors that are likely causal to each trait. In conclusion, Brain Catalog presents a one-stop shop for the genetic components of brain-related traits, potentially serving as a valuable resource for worldwide researchers to advance the understanding of how GWAS signals may contribute to the biological etiology of brain-related traits.

INTRODUCTION

Brain-related traits, including brain-related disorders, diseases and risk factors, broadly affect the common health around the world, imposing heavy burdens on patients, families, and the health system. Many complex phenotypes are related to dysfunctions in the brain, including psychiatric disorders, neurodevelopmental disorders, cognitive disorders, substance use disorders, behavioral habits, psychosocial and personality traits, and neurodegenerative diseases, among others. According to the International Classification of Diseases (ICD) system, most of these traits can be categorized as mental and behavioral disorders or diseases of the nervous system. In general, they share overlapping symptoms, pathogenesis, and common genetic risks. In this work, we generally referred to them as brain-related traits. Characterization of the genetic landscape and the manifesting tissues and cell types can provide unique insights into the underlying pathophysiological processes and the complex relationships among them.

So far, many large-scale genome-wide association studies (GWAS) have been successfully conducted, reporting hundreds of thousands of susceptibility loci associated with various brain-related traits (1–5). There are many resources available online hosting GWAS summary statistics for various traits and diseases, such as NHGRI-EBI GWAS Catalog (6), GWASdb (7), GRASP (8), PhenoScanner (9), MR-Base (10) and GWAS ATLAS (11). However, GWAS results are not readily available to pinpoint causal variants, causal genes, or the underlying mechanisms. Nearly 90% of trait-associated variants are located in non-coding regions (12) and are known to play regulatory roles. The complex linkage disequilibrium (LD) structure further complicates the identification of causal variants and genes. Thus, secondary analyses integrating multi-omics data are often necessary to

*To whom correspondence should be addressed. Tel: +86 1084097798; Email: peilin.jia@gmail.com; pjia@big.ac.cn
Correspondence may also be addressed to Zhang Zhang. Email: zhangzhang@big.ac.cn
Correspondence may also be addressed to Yiming Bao. Email: baoym@big.ac.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

deeply investigate the roles of genetic variants and the target genes they regulate.

Recently, multiple molecular quantitative trait loci (xQTL) have been generated from various platforms and studies (13) for gene expression (eQTL), gene splicing (sQTL), DNA methylation (mQTL), chromatin accessibility (caQTL), H3K27 histone acetylation (haQTL) and protein abundance (pQTL), providing valuable resources to prioritize causal variants and causal genes across tissues and cell types. However, a comprehensive, systematic resource to yield knowledge from SNPs to genes, functional tissues or cells, and putative risk factors has been lacking, especially for brain-related traits. Due to the complex comorbidity and shared genetic components among many medical and psychiatric disorders, it is of special interest to compare across many types of brain-related traits and investigate the shared and unique causal variants and genes, and the involved tissues and brain cell types. Prioritizing causal variants and identifying their targeting genes is a daunting job that involves integrating multi-omics datasets across multiple tissues and cell types. There have already been works attempting to systematically fine-map candidate susceptibility genes for published GWAS summary statistics, such as CAUSALdb (14) and webTWAS (15). However, for brain-related traits, a resource has been lacking in integrating all relevant functional annotation and multi-omics datasets and implementing different methods to infer loci, genes, tissues, and cell types that mediate the genetic mechanisms of disease and the relationships with other diseases or risk factors.

In this study, we developed a comprehensive resource, the Brain Catalog, available at https://ngdc.cncb.ac.cn/braincatalog. Brain Catalog curated the majority of published GWAS summary statistics for brain-related traits from multiple consortia and studies. It presents inference of causal variants, causal genes, functional tissues or cell types, and potential risk factors from variant-, gene- and trait-levels with multiple functional annotation tracks.

MATERIALS AND METHODS

GWAS summary statistics data curation

We downloaded GWAS summary statistics for brain-related traits that were publicly available from online resources. The types of traits that were considered in the current version were mainly searched using the terms from two groups as defined by ICD-10: Chapter V Mental and behavioral disorders (F00–F99) and Chapter VI Diseases of the nervous system (G00–G99). Other traits were collected based on literature citations, experiences, and expertise knowledge such as the brain imaging data. So far, we have collected a total of 517 GWAS summary statistics datasets that can be categorized into two groups according to the phenotype definition: one with canonical phenotype measurements such as binary, categorical or continuous phenotypes (termed the non-image group) and the other with image-based phenotypes. For the non-image group of phenotypes, we downloaded data mainly from Psychiatric Genomics Consortium (PGC) (16), NHGRI-EBI GWAS Catalog (6), Neale Lab UKBB v3 (17), Complex Trait Genetics (CTGlab) (https://ctg.cncr.nl/), the International Multiple Sclerosis Genetics Consortium (IMSGC) (18), the MEGASTROKE (19), the International Cannabis Consortium (ICC) (20,21), the GWAS & Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) (22), the Social Sciences Genetic Association Consortium (SSGAC) (23,24), the Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE) (25), the Genetics of Personality Consortium (GPC) (26) and The Sleep Disorder Knowledge Portal (SDKP) (27). For human brain imaging traits, we downloaded GWAS summary statistics data from Biostatistics and Imaging Genomics analysis lab (BIG-KP) (28) and the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) (29). Redundant datasets were removed. For each dataset, we extracted the sample size, ancestry, and other basic information. Considering that the majority of the reference panels and xQTL annotations used for post-GWAS analyses were from European ancestry, we only included the GWAS data from European ancestry in the current version of Brain Catalog.

Quality control of GWAS summary statistics

The downloaded GWAS summary statistics data were saved in various formats depending on the platforms or preprocessing procedures. We implemented a series of quality control steps to reformat the raw data. Firstly, all GWAS datasets were converted to GRCh37 (hg19) coordinates. Secondly, alleles of each variant were mapped to the 1000 Genomes Project (1KGP) reference panel to keep consistency. In cases where the effect size (beta) is missing, we imputed beta by using other statistical data such as the standard error (SE), z-score, and *P*-value. For a small subset (36 cases, 6.96%) of the downloaded data, only the *P*-values were available and there was no possible way to impute other statistics such as the effect size and SE and thus, we only applied a part of the methods wherever applicable. Thirdly, allele frequency was filled in using the European population from 1KGP (phase 3). If an imputation score (INFO) was available, variants with INFO <0.9 were excluded. Finally, we performed other quality controls using MungeSumstats (30) based on dbSNP144 (31), such as inferring missing SNP ID, chromosome (CHR), or base pair position (BP), removing non-biallelic SNPs, and harmonizing GWAS summary statistics to the reference genome.

xQTL resources

We collected a total of 58 xQTL datasets belonging to six different types of QTL annotation data, especially for brain and blood. These included 27 expression QTL (eQTL) datasets, 22 sQTL (splicing QTL) datasets, 5 mQTL (methylation) datasets, 2 pQTL (protein QTL) datasets, 1 caQTL (chromatin accessibility) dataset and 1 haQTL (H3K27 histone acetylation) dataset. We included only the *cis*-xQTL for the following analyses. More details of these datasets, such as the origin tissues, QTL types, and sources, are presented in Supplementary Table S1. Notably, due to the limited resources, we did not include trans-QTL datasets in the current version of Brain Catalog.

SNP-based heritability, functional partition of heritability and genetic correlation

Three types of analyses were conducted for heritability estimation by using two methods, i.e. LD score regression (LDSC, v1.0.1) (32) and Linkage-Disequilibrium Adjusted Kinships (LDAK, v5.2) (33). For these analyses, all GWAS summary statistics were filtered using HapMap3 SNPs and required to have frequency (FRQ) >0.01. Large-effect loci (e.g. those in the MHC region) were excluded from all analyses. First, we estimated the SNP-based heritability for each trait by using LDSC and BLD-LDAK. Second, the heritability enrichment of a category of SNPs is estimated using LDAK. We used 26 binary categorical annotations for SNPs, including super-enhancers, conserved regions, and so on. Third, genetic correlations were estimated by using both LDSC and LDAK (the BLD-Thin model). To provide a reliable estimation of genetic correlations, the traits were required to have (i) SNP-based heritability between 0 and 1, (ii) *P*-value <0.05, (iii) heritability (H^2) *z*-score >1.5, (iv) mean chi-square >1.02 and (v) the intercept between 0.9 and 1.1.

Cell and tissue functional enrichment

We used two methods to identify functional tissues or cell types for the traits in the Brain Catalog: GARFIELD v2 (34) and stratified LD score regression (S-LDSC) (35). GARFIELD assessed whether trait-related variants, as defined using different significance thresholds, were enriched in different regulatory or functional annotations using a logistic regression mode. GARFIELD provided 1005 annotated features collected from ENCODE, GENCODE and the Roadmap Epigenomics project. These features included genic annotations, histone modifications, chromatin status, transcription factor binding sites, and open chromatin data (such as FAIRE, and DNase I hypersensitive site hotspots) and were obtained using 424 cell types and 55 tissues. We also used S-LDSC and single cell ATAC-seq (scATAC-seq) data from the human cortex (36) to evaluate whether these brain-related traits were enriched in cell-type specific open chromatin regions.

Fine-mapping with multiple annotation tracks

To perform fine-mapping on brain-related traits, we first extracted genome-wide significant SNPs (*P*-value $< 5 \times 10^{-8}$) as index SNPs. For each index SNP, we defined a genetic locus as a ±500 kb window centered around it, with low-frequency SNPs (i.e. FRQ < 0.01) filtered in the region. Next, statistical fine-mapping was conducted for each locus using four strategies including ABF, FINEMAP, SuSiE and PolyFun + SuSiE, all of which have been integrated in the R package *echolocatoR* (37). By estimating the posterior probability (PP) of each SNP being a causal SNP, these fine-mapping methods generate the following key SNP sets: (i) 95% credible sets ($CS_{95\%}$) for each causal signal by each method, (ii) union credible set of SNPs (UCS), which was defined as the union of all method-specific $CS_{95\%}$ and (iii) consensus SNPs defined as those identified by at least two methods. We set the (maximum) number of causal SNPs to five for all fine-mapping tools, except for ABF with only one assumed causal SNP. Additionally, all loci were annotated with multiple cell type-specific epigenomic annotations from Nott *et al.* (38), Corces *et al.* (39) and eXploring Genomic Relations (XGR) (40) including cell type-specific epigenomic peaks (scATAC, ATAC, H3K27ac, H3K4me), cell type-specific regulatory regions (enhancers, promoters), cell type-specific interactome anchors (proximity ligation-assisted ChIP-Seq), bulk brain epigenomic peaks (ATAC), and bulk brain interactome anchors (HiChIP_FitHiChIP).

Prioritizing causal variants and causal genes with xQTL

We prioritized causal variants and causal genes by combining GWAS summary statistics with six cis-xQTLs (eQTL, sQTL, mQTL, pQTL, haQTL, caQTL) from multiple tissues using three methods, which were transcriptome-wide association analysis (TWAS) (41–43), Bayesian colocalization analysis (COLOC) (44), and summary data-based mendelian randomization analysis (SMR) (45). Specifically, for the TWAS method, we utilized three widely-applied methods to avoid missing of candidate risk genes: S-PrediXcan, unified test for molecular signatures (UTMOST), and joint-tissue imputation (JTI). We used the prediction models trained on GTEx v8 available at the PredictDB website (http://predictdb.org) and GitHub (https://github.com/gamazonlab/MR-JTI). For colocalization analysis, we used the R package *coloc* (44) to assess the probability of the same variant being responsible between brain-related traits and xQTLs. The recommended combination thresholds (PP4 ≥ 0.75 and PP4/PP3 ≥ 3) were employed to define a causal role for the tested gene. Finally, SMR was performed to test the association between an exposure (here a quantitative trait such as gene expression) and an outcome (brain-related trait) based on the Mendelian randomization framework using a genetic variant as the instrument variable (IV). We also applied the HEIDI test to distinguish pleiotropy from linkage with a threshold of 0.05. The Benjamini-Hochberg method correction was used in each TWAS or SMR analysis to control the false discovery rate (FDR < 0.05).

Risk factors

Mendelian randomization (MR) analyses were performed to assess the causal effect between exposures and outcome traits using the R package *TwoSampleMR* (10). For each exposure, significant SNPs (defined as those with $P < 5 \times 10^{-8}$, $P < 1 \times 10^{-6}$, or $P < 1 \times 10^{-5}$ whichever applicable depending on the power of the original GWAS) served as the Instrument Variables (IVs). Then, we performed a clumping procedure using the 1KGP European samples as LD reference panel and requiring $r^2 < 0.05$ in a 1000-kb window. Finally, we applied six different MR methods for a comprehensive investigation: *Wald ratio*, *MR Egger*, *Weighted median*, *Inverse variance weighted*, *Simple mode* and *Weighted mode*. A series of sensitivity analyses were also carried out, including the heterogeneity test, horizontal pleiotropy test and leave-one-out analysis.

Database design and updates

Brain Catalog is hosted by a virtual machine with CentOS 7.9. The backend RESTful web service was constructed with Java Spring Boot (https://spring.io/projects/spring-boot) framework and MySQL (https://www.mysql.com) as the database engine. The web frontend interface was developed using the React (https://reactjs.org) and UMI (https://umijs.org) frameworks, which are scalable enterprise-class frontend frameworks containing a complete life cycle plugin system that can be extended to execute more complex functionality. Ant Design (https://ant.design) is used as the UI library that contains a set of high-quality components for building rich, interactive user interfaces. In addition, the interactive visual charts were implemented by using different libraries such as HighCharts (https://www.highcharts.com), ECharts (https://echarts.apache.org), PlotlyJS (https://plotly.com/javascript) and UpSet.js (https://upset.js.org).

All analyses have been implemented by customized scripts to enable automatic and efficient processing of large-scale datasets. Considering the rapid development of GWAS data, where new releases with large sample sizes and novel discoveries are expected in the coming years, we will conduct regular updates of the database every six months. New GWAS data will be searched and processed on a regular basis and the corresponding analysis results will be updated in the database every half a year. In addition to the updates of GWAS data, new annotation datasets such as various QTL data will also be included to expand the analysis categories and to provide candidate causal variants and factors for each trait.

DATABASE CONTENT AND USAGE

Schematic overview

Brain Catalog currently contains 517 brain-related traits from various consortia and studies, 58 xQTL datasets belonging to six types of QTL mapping covering 21 tissues, 436 cell type annotation data, and results from 22 post-GWAS methods. The traits covered as many as possible traits that were generally related to the brain, including psychiatric disorders, neurodegenerative diseases, brain imaging, behavioral habits, substance use disorders, cognitive function, personality traits, psychosocial characteristics, and others. We conducted a series of post-GWAS analyses to delineate the genetic components of each trait and the genetic correlations among traits. Briefly, at the genetic variant level, we applied GARFIELD and S-LDSC to evaluate the enrichment of associated-trait loci in functional annotations across multiple tissues and cell types. We also prioritized causal variants and causal genes using nine widely-used methods (four fine-mapping, three TWAS, SMR and COLOC) and six types of xQTLs focusing on brain and blood. This part of the analyses can provide insights into the regulatory mechanisms as to how a candidate genetic variant may mediate the quantitative trait changes (such as gene expression changes) and eventually contribute to the disease outcome. At the heritability level, we estimated SNP-based heritability, functional partition of heritability enrichment, and cross-trait genetic correlation using both LDSC and LDAK. Finally, we performed the 2SMR analysis to estimate the causal effect among all traits using six methods, followed by sensitivity analyses. Although in literature, risk factors, such as smoking, sleeping duration, and alcohol consumption, are often used as the exposure variable and disorders as the outcome such as schizophrenia, Parkinson's disease and autism, we systematically conducted 2SMR to pair-wise traits in our database without distinguishing risk factors and disorders. Thus, the results presented in Brain Catalog have well-controlled biases likely introduced by *a priori* selection. A schematic overview of Brain Catalog is illustrated in Figure 1.

Web interface

Brain Catalog provides a user-friendly interface for users to access information by searching, browsing, and downloading. The search function allows users to quickly search by using keywords for traits, SNPs, or genes of interest (Figure 2A). The browse menu provides quick access to all major categories of analysis, such as the heritability analysis, two sample MR, SMR, colocalization, and fine-mapping. Brain Catalog organizes the results by traits and further categorized the contents by five analysis themes (11 modules) including (i) tissues or cell type functional enrichment, (ii) heritability estimation, heritability enrichment and genetic correlation, (iii) fine-mapping with multiple annotation tracks, (iv) prioritizing causal genes with xQTLs, (v) MR analysis of putative risk factor (Figure 2C). Thus, the best strategy is to start with the trait overview which displays summaries of all 517 brain-related traits, such as trait names, type of phenotypes, sample sizes, and basic information of each study (Figure 2A), and then select the trait of interest to access the overall analysis modules (Figure 2B, C).

Almost all units provide interactive analysis, online visualization, and download options. For example, for the heritability analyses, we used a barplot to show the estimated SNP heritability by both LDSC and LDAK, a dot plot accompanied with a table to show the heritability enrichment results, and a trait-trait network to show the genetic correlations. All figures and tables are interactive and users can select one or more traits of interest to have a zoom-in visualization (Figure 2D). For the analyses where multiple methods were implemented, such as the TWAS analyses (involving three methods) and the COLOC/SMR analyses (involving 58 QTL datasets), we use an UpSet plot and a Venn diagram to show the overlapping results, a heatmap to show the SMR results for each xQTL in the corresponding tissue, and a local Manhattan plot to show the colocalization results for each locus (Figure 2E).

Amyotrophic lateral sclerosis (ALS) as an example application

Below we used the ALS GWAS data reported by van Rheenen *et al.* (46) as an example to demonstrate the usage of Brain Catalog.

SNP-base heritability, heritability enrichment and genetic correlation. By using two methods, SNP-based heritability for ALS was estimated to be 3.8% (SE = 0.0044,

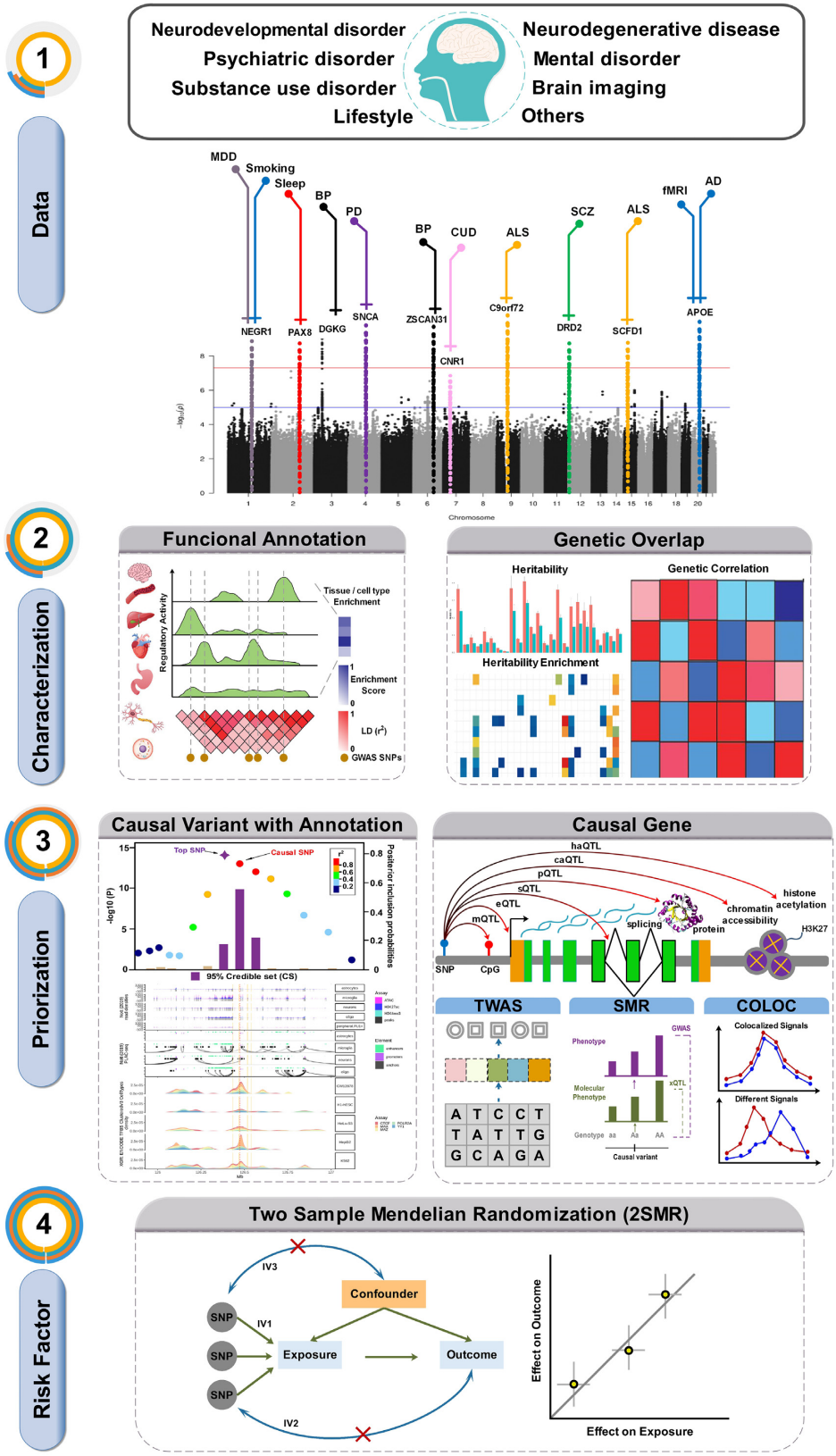

Figure 1. Schematic overview of the Brain Catalog. It consists of four main components. (1) Data resources: brain-related traits collection from different consortia and studies. (2) Characterization: cell and tissue functional enrichment, SNP-based heritability, heritability enrichment, and genetic correlation. (3) Prioritization of causal variants and causal genes: fine-mapping with multiple annotation tracks and identify causal genes with six xQTLs. (4) Risk factors: MR estimation of the causal effect between brain-related traits using six methods.

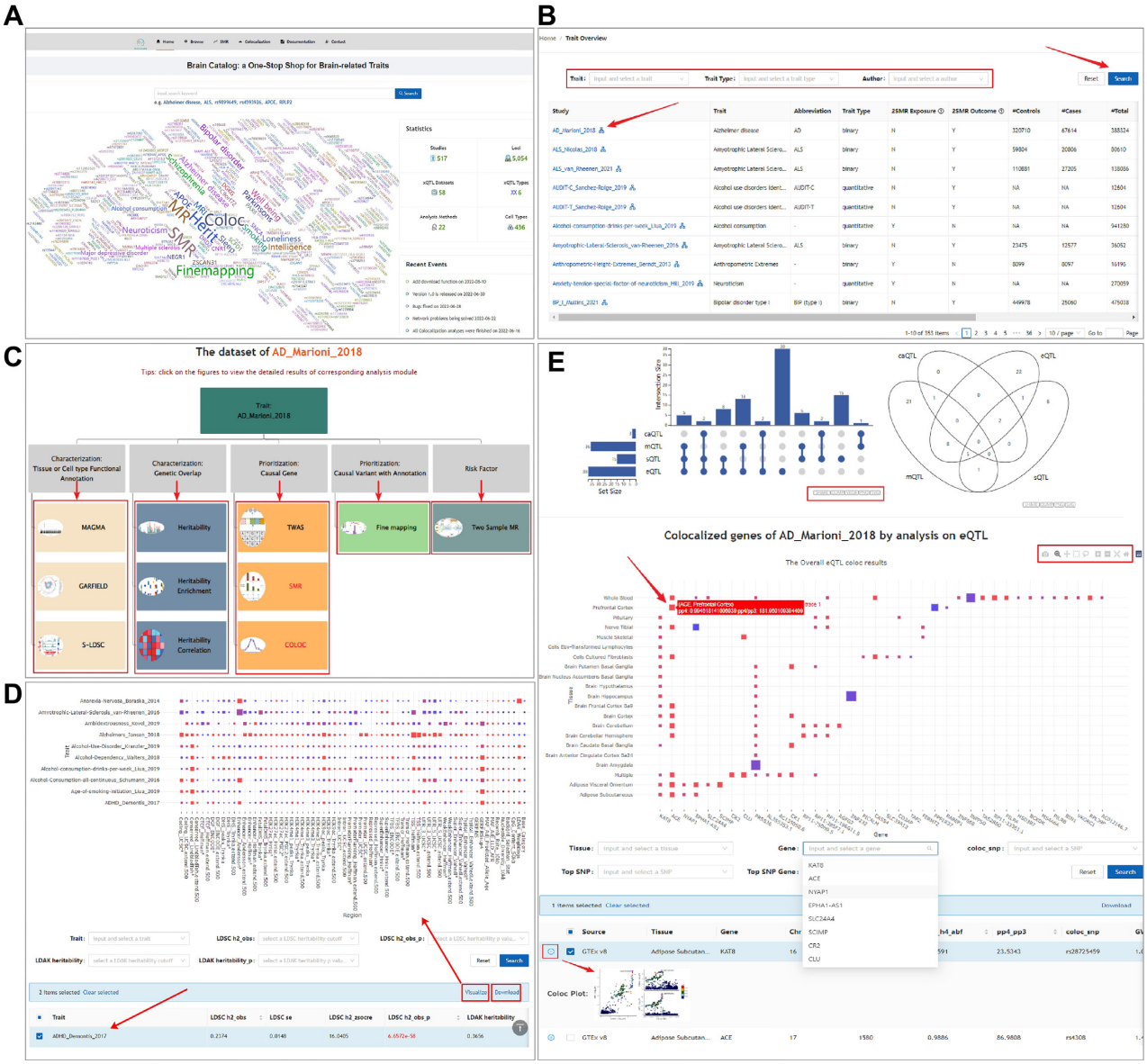

Figure 2. Screenshots of web pages for the Brain Catalog. (**A**) The home page allows for quick research for traits, SNPs, genes, or analysis topics. (**B**) An overview table for all traits is provided in the browse menu. (**C**) Results of multiple analyses for the dataset AD_Marioni_2018. The red boxes highlight the analysis modules available for selection. (**D**) Interactive figures and tables in the heritability module to show the results from multiple analysis methods. (**E**) Results of colocalization with xQTL. The color and size of the grid are proportional to the probability and intensity of colocation signals. A local Manhattan plot is also provided for visualization.

$P = 3.9 \times 10^{-18}$) by LDSC and 8.2% (SE = 0.0076, $P = 1.02 \times 10^{-26}$) by LDAK, respectively (Supplementary Figure S1A), consistent with previous studies (47). Using 26 binary category annotations, partitioned LDAK heritability analysis showed significant enrichment for SNPs located in 9 different categories, especially for super-enhancer (Enrichment OR = 1.98, $P = 1.38 \times 10^{-7}$) (Supplementary Table S2). Moreover, genetic correlation analysis by both LDSC and LDAK suggested that ALS shared genetic pathogenicity with multiple diseases and traits, including Alzheimer's disease (LDSC: $r_g = 0.45$, $P = 1.4 \times 10^{-3}$; LDAK: $r_g = 0.46$, $P = 2.27 \times 10^{-6}$), educational attainment (LDSC: $r_g = -0.17$, $P = 2.73 \times 10^{-6}$; LDAK: $r_g = -0.17$, $P = 2.63 \times 10^{-6}$), and alcohol use disorder (LDSC: $r_g = 0.22$, $P = 6 \times 10^{-4}$; LDAK: $r_g = 0.17$, $P = 0.01$) (Supplementary Figure S1B).

Functional enrichment and annotation. Functional enrichment analysis was conducted using S-LDSC and GARFIELD for ALS-associated loci defined using five different thresholds: P-value $< 1 \times 10^{-8}$, P-value $< 1 \times 10^{-7}$, P-value $< 1 \times 10^{-6}$, P-value $< 1 \times 10^{-5}$ and P-value $< 1 \times 10^{-4}$. Using GARFIELD, we found the largest fold enrichment in the urothelium (OR = 10.79, P-value $= 4.94 \times 10^{-6}$) for the locus set defined using P-value $< 1 \times 10^{-7}$, which may be related to the comorbidities such as urinary infections or incontinence that caused death in ALS patients (48,49) (Figure 3A). Notably, S-

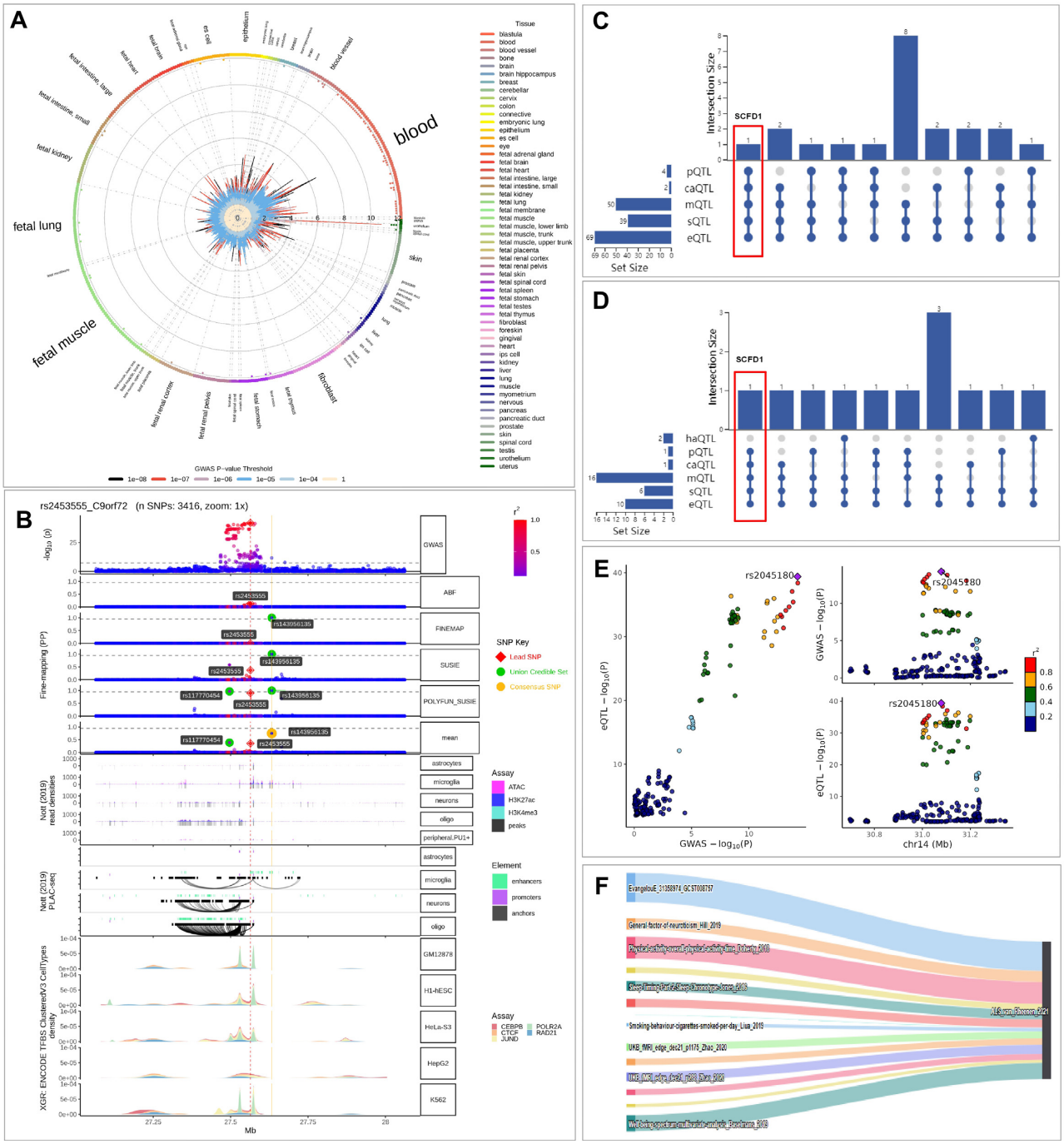

Figure 3. The post-GWAS analysis for ALS. (**A**) Enrichment results by GARFIELD of ALS-associated variants in DNase I Hypersensitive sites hotspot regions from ENCODE and Roadmap Epigenomics data. The radial plot shows the enrichment (OR) in each of the 424 cell types sorted by tissue on the outside edge of the circle. The size of the tissue label was proportional to the number of cell types. Enrichment analysis was performed using five significance thresholds (1×10^{-8}, 1×10^{-7}, 1×10^{-6}, 1×10^{-5}, 1×10^{-4}), which were shown by different colors. The dots on the inner side of the outer circle represented the significance level of enrichment analysis for thresholds 1×10^{-5}, 1×10^{-6}, 1×10^{-7} and 1×10^{-8} in the direction from outside to inside. (**B**) Fine-mapping of the C9orf72 locus. The regional Manhattan plot was shown in the top panel. The top SNP was labeled as a diamond, and the color indicated LD (r^2) between each SNP and the top SNP. Below the Manhattan plot, we show four tracks for four fine-mapping methods (labeled on the right of the panel: ABF, FINEMAP, SuSiE, and PolyFun + SuSiE,) and a track (labeled as 'mean' on the right of the panel) showing the mean PPs of all four methods. In all these tracks, the y-axis showed the PP value of SNPs (ranging between 0–1). The union credible set of SNPs (UCS) were labeled in green and consensus SNPs were labeled in gold. Three epigenetic annotation tracks were shown at the bottom. (**C**) The summary of significant genes identified by SMR with xQTLs. (**D**) The summary of colocalized genes by COLOC with xQTLs. (**E**) Colocalization of the G2E3 locus with eQTL in the prefrontal cortex (PsychENCODE). The color indicated the LD degree between the candidate causal SNP (in this case, rs2045180, labeled as a purple diamond) and other SNPs. (**F**) MR results between several risk factors and ALS as an outcome. The width of the line represented the effect size of the exposure on ALS. Detailed results are available in Supplementary Table S7.

LDSC analysis did not identify significantly enriched functional cells for ALS, further demonstrating the necessity to use multiple methods for each type of analysis.

Fine-mapping with multiple annotation tracks. By applying four fine-mapping methods (ABF, FINEMAP, SuSiE and PolyFun + SuSiE) for 13 risk loci in ALS ($P < 5 \times 10^{-8}$), we identified 5 consensus SNPs, each identified by at least two methods. They were rs116900480 in the locus rs116900480_ATP23 (mean posterior probability or PP = 0.95), rs12608932 in rs12608932_UNC13A (mean PP = 0.92), rs143956135 in rs2453555_C9orf72 (mean PP = 0.75), rs73440960 in rs535957039_LINGO2 (mean PP = 0.75), rs75087725 in rs75087725_CFAP410 (mean PP = 0.99) (Supplementary Table S3). Further COLOC analyses showed rs73440960 was colocalized with sQTL in brain caudate basal ganglia with its nearest gene *C9orf72* but no significant colocalization signal was found with any xQTL for the other consensus SNPs. Interestingly, in the rs2453555_C9orf72 locus, although the candidate consensus SNP was determined as rs143956135, a nearby SNP, rs2453555, was also reported by one method (PolyFun + SuSiE) with a high posterior probability (PP = 0.9). Furthermore, COLOC analysis revealed that rs2453555 showed significant colocalization signals with eQTL (PP4 = 0.81, PP4/PP3 = 5.8) or sQTL in brain (PP4 = 0.97, PP4/PP3 = 30.7) (Figure 3B, Supplementary Table S6), which was in line with our previous study (50).

Prioritizing causal variants and causal genes with xQTLs. By using three TWAS methods (S-PrediXcan, UTMOST and JTI), we identified a total of 102 genes significantly associated with ALS by at least two methods (FDR < 0.05), of which 19 were reported in recent GWAS and post-GWAS studies (46,50,51) (Supplementary Table S4), leaving a large number of newly identified candidate genes for ALS. By using the SMR analysis, we tested the exposure measurements from six types of cis-xQTLs and identified a total of 84 genes with significant evidence that likely mediated the genetic associations (FDR < 0.05 and HEIDI > 0.05). As expected, the maximum number of genes ($n = 43$) were found with eQTL datasets (Supplementary Table S5). A novel finding of particular interest was the gene *SCFD1*, which was identified by five kinds of xQTLs (Figure 3C, Supplementary Table S5), suggesting *SCFD1* was likely involved in multiple functions underlying ALS pathogenesis.

Colocalization analysis was conducted for each of the significant ALS-associated SNPs ($P < 5 \times 10^{-8}$) using 58 xQTL datasets, respectively. In total, we identified 21 candidate genes with significant evidence in at least one tissue (PP4 > 0.75 and PP4/PP3 > 3) (Figure 3D, Supplementary Table S6). The largest number of signals ($n = 16$) was found with mQTL datasets (Supplementary Table S6). The strongest signal was identified for the gene *G2E3* (PP4 = 0.99 and PP4/PP3 = 400.5) in the prefrontal cortex eQTL dataset (Figure 3E, Supplementary Table S6). Interestingly, consistent with the finding in SMR, *SCFD1* was colocalized in five kinds of xQTLs, further highlighting its importance (Figure 3D, Supplementary Table S6).

Risk factors causally related to ALS. A comprehensive MR analysis with ALS as the outcome showed that several exposures were identified as causal risk factors for ALS by at least one method with weak effect sizes (MR $P < 0.05$, horizontal pleiotropy P-value > 0.05), including extreme BMI, alcohol consumption, sleeping, neuroticism, physical activity, educational attainment, smoking, well-being, and brain imaging (Figure 3F, Supplementary Table S7). Among these factors, neuroticism, educational attainment, well-being, and physical activity time showed a negative effect on ALS, which was consistent with previous reports (46,52). A positive effect of alcohol consumption on ALS was identified in our analyses, but previous studies had reported both positive effects and negative effects for alcohol consumption (53). Several results suggested that hippocampal volume and dentate gyrus volume had a negative effect on ALS (Supplementary Table S7). Interestingly, in neuroimaging traits, eight amplitude traits (node) reflecting regional spontaneous neuronal activity and two functional connectivity traits (edge) quantifying the interregional coactivity showed a positive effect on ALS, and two functional connectivity traits showed a negative effect on ALS (Supplementary Table S7). These results highlighted a complex mechanism for the relationship between brain images and ALS.

DISCUSSION AND FUTURE DIRECTIONS

We developed Brain Catalog, a comprehensive resource aiming at delineating the genetic components underlying brain-related traits from various aspects. Brain Catalog integrated GWAS summary statistics and multi-omics datasets by using multiple cutting-edge analytical approaches. The rich results from these analyses provided unique insights into the underlying disease mechanisms as well as genetic correlations among traits. To the best of our knowledge, Brain Catalog is currently the largest and most comprehensive database for brain-related traits covering both disorders/diseases and risk factors, allowing cross-methods, cross-QTLs, and cross-traits comparison.

Although brain-related traits may refer to a wide range of phenotypes, we focus on those with reported genetic components and have been studied by GWAS. Currently, the traits included in Brain Catalog nearly all belonged to the mental and behavioral disorders and diseases of the nervous system. For each trait, Brain Catalog provides quite comprehensive information to allow in-depth single-trait analysis and cross-trait analysis. First, it implements multiple methods to prioritize causal variants and causal genes with a wide range of xQTLs. Second, it provides an estimation of SNP heritability, functional partition of heritability enrichment, and cross-trait genetic correlation. Third, it provides the risk factor landscape for each brain-related trait. These results collectively formed a valuable resource for the genetic landscape of most traits and also enabled cross-trait integration to study comorbidity and shared heritability.

Although the current version of Brain Catalog includes very comprehensive datasets and annotations, there are several limitations that need to be addressed in future studies. First, given that all analyses were based on GWAS summary statistics, we were not able to assess the possible overlapping of samples among studies, which could reduce the

power of the analysis and introduce potential biases, especially for the genetic correlation and MR analyses. In addition, the reference panel we used, i.e. 1KGP, may not reflect the actual LD pattern of the corresponding populations, although replication between different studies could help to reduce biases. For example, we collected 5 GWAS summary statistics for schizophrenia from different studies. Results from these datasets would provide cross-study validation of each other. Second, the current version of our database only contains European populations and thus, the results may not be generalizable to other populations. Third, although we selected 21 major tissues for SMR and COLOC, other possible pathogenic tissues were not included. Finally, due to the limited resources of trans-xQTLs, we only included cis-QTL datasets in our analysis, which may miss causal variants or causal genes that function in *trans*. In future, more xQTL datasets, especially those based on single-cell data, will be integrated to fully prioritize causal variants and causal genes to provide more functional information.

In conclusion, Brain Catalog provides a one-stop shop for the genetic landscape of brain-related traits. We expect it to serve the broad research area to explore the genetic mechanisms underlying diseases.

DATA AVAILABILITY

All processed GWAS summary statistics can be downloaded in the Brain Catalog (https://ngdc.cncb.ac.cn/braincatalog). xQTL datasets are publicly available in GTEx Portal (https://gtexportal.org/), eQTLGen consortium (https://www.eqtlgen.org/), SMR (https://yanglab.westlake.edu.cn/software/smr/), Bran xQTL server (http://mostafavilab.stat.ubc.ca/xqtl/) and studies.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank all the platforms that provided GWAS summary statistics, including PGC, GWAS Catalog, ENIGMA, BIG-KP, CTGlab, UKBB, ICC, GSCAN, SSGAC, GPC, Broad ABC, MEGASTROKE, IMSGC, CCACE, EGG. We thank NGDC for its support in providing computing resources in data processing and database construction. We are grateful for the technical support provided by the Laboratory of Precision Health.

FUNDING

Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38010400, XDB38010401]. Funding for open access charge: Strategic Priority Research Program of the Chinese Academy of Sciences.
Conflict of interest statement. None declared.

REFERENCES

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.
2. Wray,N.R., Ripke,S., Mattheisen,M., Trzaskowski,M., Byrne,E.M., Abdellaoui,A., Adams,M.J., Agerbo,E., Air,T.M., Andlauer,T.M.F. *et al.* (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.*, **50**, 668–681.
3. Nalls,M.A., Blauwendraat,C., Vallerga,C.L., Heilbron,K., Bandres-Ciga,S., Chang,D., Tan,M., Kia,D.A., Noyce,A.J., Xue,A. *et al.* (2019) Identification of novel risk loci, causal insights, and heritable risk for parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.*, **18**, 1091–1102.
4. Bellenguez,C., Kucukali,F., Jansen,I.E., Kleineidam,L., Moreno-Grau,S., Amin,N., Naj,A.C., Campos-Martin,R., Grenier-Boley,B., Andrade,V. *et al.* (2022) New insights into the genetic etiology of alzheimer's disease and related dementias. *Nat. Genet.*, **54**, 412–436.
5. Visscher,P.M., Wray,N.R., Zhang,Q., Sklar,P., McCarthy,M.I., Brown,M.A. and Yang,J. (2017) 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.*, **101**, 5–22.
6. 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.
7. Li,M.J., Liu,Z., Wang,P., Wong,M.P., Nelson,M.R., Kocher,J.P., Yeager,M., Sham,P.C., Chanock,S.J., Xia,Z. *et al.* (2016) GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, **44**, D869–D876.
8. Eicher,J.D., Landowski,C., Stackhouse,B., Sloan,A., Chen,W., Jensen,N., Lien,J.P., Leslie,R. and Johnson,A.D. (2015) GRASP v2.0: an update on the genome-wide repository of associations between SNPs and phenotypes. *Nucleic Acids Res.*, **43**, D799–D804.
9. Kamat,M.A., Blackshaw,J.A., Young,R., Surendran,P., Burgess,S., Danesh,J., Butterworth,A.S. and Staley,J.R. (2019) PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics*, **35**, 4851–4853.
10. 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.
11. Watanabe,K., Stringer,S., Frei,O., Umicevic 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.
12. Farh,K.K., Marson,A., Zhu,J., Kleinewietfeld,M., Housley,W.J., Beik,S., Shoresh,N., Whitton,H., Ryan,R.J., Shishkin,A.A. *et al.* (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, **518**, 337–343.
13. Liu,L., Zhang,Y., Niu,G., Li,Q., Li,Z., Zhu,T., Feng,C., Liu,X., Zhang,Y., Xu,T. *et al.* (2022) BrainBase: a curated knowledgebase for brain diseases. *Nucleic Acids Res.*, **50**, D1131–D1138.
14. Wang,J., Huang,D., Zhou,Y., Yao,H., Liu,H., Zhai,S., Wu,C., Zheng,Z., Zhao,K., Wang,Z. *et al.* (2020) CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. *Nucleic Acids Res.*, **48**, D807–D816.
15. Cao,C., Wang,J., Kwok,D., Cui,F., Zhang,Z., Zhao,D., Li,M.J. and Zou,Q. (2022) webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. *Nucleic Acids Res.*, **50**, D1123–D1130.
16. Sullivan,P.F., Agrawal,A., Bulik,C.M., Andreassen,O.A., Borglum,A.D., Breen,G., Cichon,S., Edenberg,H.J., Faraone,S.V., Gelernter,J. *et al.* (2018) Psychiatric genomics: an update and an agenda. *Am. J. Psychiatry*, **175**, 15–27.
17. Sudlow,C., Gallacher,J., Allen,N., Beral,V., Burton,P., Danesh,J., Downey,P., Elliott,P., Green,J., Landray,M. *et al.* (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.*, **12**, e1001779.
18. International Multiple Sclerosis Genetics, C. (2019) Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*, **365**, eaav7188.
19. Malik,R., Chauhan,G., Traylor,M., Sargurupremraj,M., Okada,Y., Mishra,A., Rutten-Jacobs,L., Giese,A.K., van der Laan,S.W., Gretarsdottir,S. *et al.* (2018) Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.*, **50**, 524–537.

20. Stringer,S., Minica,C.C., Verweij,K.J., Mbarek,H., Bernard,M., Derringer,J., van Eijk,K.R., Isen,J.D., Loukola,A., Maciejewski,D.F. *et al.* (2016) Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the international cannabis consortium. *Transl Psychiatry*, **6**, e769.
21. Pasman,J.A., Verweij,K.J.H., Gerring,Z., Stringer,S., Sanchez-Roige,S., Treur,J.L., Abdellaoui,A., Nivard,M.G., Baselmans,B.M.L., Ong,J.S. *et al.* (2018) GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.*, **21**, 1161–1170.
22. Liu,M., Jiang,Y., Wedow,R., Li,Y., Brazel,D.M., Chen,F., Datta,G., Davila-Velderrain,J., McGuire,D., Tian,C. *et al.* (2019) Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.*, **51**, 237–244.
23. Karlsson Linner,R., Biroli,P., Kong,E., Meddens,S.F.W., Wedow,R., Fontana,M.A., Lebreton,M., Tino,S.P., Abdellaoui,A., Hammerschlag,A.R. *et al.* (2019) Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.*, **51**, 245–257.
24. Meddens,S.F.W., de Vlaming,R., Bowers,P., Burik,C.A.P., Linner,R.K., Lee,C., Okbay,A., Turley,P., Rietveld,C.A., Fontana,M.A. *et al.* (2021) Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *Mol. Psychiatry*, **26**, 2056–2069.
25. Deary,V., Hagenaars,S.P., Harris,S.E., Hill,W.D., Davies,G., Liewald,D.C.M. and International Consortium for Blood Pressure, G.International Consortium for Blood Pressure, G., Aging,C.C., Longevity,G., Group,C.C.I. *et al.* (2018) Genetic contributions to self-reported tiredness. *Mol. Psychiatry*, **23**, 609–620.
26. van den Berg,S.M., de Moor,M.H., Verweij,K.J., Krueger,R.F., Luciano,M., Arias Vasquez,A., Matteson,L.K., Derringer,J., Esko,T., Amin,N. *et al.* (2016) Meta-analysis of genome-wide association studies for extraversion: findings from the genetics of personality consortium. *Behav. Genet.*, **46**, 170–182.
27. Cade,B.E., Lee,J., Sofer,T., Wang,H., Zhang,M., Chen,H., Gharib,S.A., Gottlieb,D.J., Guo,X., Lane,J.M. *et al.* (2021) Whole-genome association analyses of sleep-disordered breathing phenotypes in the NHLBI TOPMed program. *Genome Med*, **13**, 136.
28. Zhao,B., Li,T., Smith,S.M., Xiong,D., Wang,X., Yang,Y., Luo,T., Zhu,Z., Shan,Y., Matoba,N. *et al.* (2022) Common variants contribute to intrinsic human brain functional networks. *Nat. Genet.*, **54**, 508–517.
29. Brouwer,R.M., Klein,M., Grasby,K.L., Schnack,H.G., Jahanshad,N., Teeuw,J., Thomopoulos,S.I., Sprooten,E., Franz,C.E., Gogtay,N. *et al.* (2022) Genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nat. Neurosci.*, **25**, 421–432.
30. 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.
31. Pagès,H. (2017) SNPlocs.Hsapiens.dbSNP144.GRCh37: SNP locations for Homo sapiens (dbSNP Build 144).
32. Bulik-Sullivan,B.K., Loh,P.R., Finucane,H.K., Ripke,S., Yang,J., Patterson,N., Daly,M.J., Price,A.L., Neale,B.M. and Schizophrenia Working Group of the Psychiatric Genomics ConsortiumSchizophrenia Working Group of the Psychiatric Genomics Consortium (2015) LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.
33. Speed,D., Holmes,J. and Balding,D.J. (2020) Evaluating and improving heritability models using summary statistics. *Nat. Genet.*, **52**, 458–462.
34. Iotchkova,V., Ritchie,G.R.S., Geihs,M., Morganella,S., Min,J.L., Walter,K., Timpson,N.J., Consortium,U.K., Dunham,I., Birney,E. *et al.* (2019) GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.*, **51**, 343–353.
35. 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.
36. Ziffra,R.S., Kim,C.N., Ross,J.M., Wilfert,A., Turner,T.N., Haeussler,M., Casella,A.M., Przytycki,P.F., Keough,K.C., Shin,D. *et al.* (2021) Single-cell epigenomics reveals mechanisms of human cortical development. *Nature*, **598**, 205–213.
37. Schilder,B.M., Humphrey,J. and Raj,T. (2021) echolocatoR: an automated end-to-end statistical and functional genomic fine-mapping pipeline. *Bioinformatics*, **38**, 536–539.
38. Nott,A., Holtman,I.R., Coufal,N.G., Schlachetzki,J.C.M., Yu,M., Hu,R., Han,C.Z., Pena,M., Xiao,J., Wu,Y. *et al.* (2019) Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*, **366**, 1134–1139.
39. Corces,M.R., Shcherbina,A., Kundu,S., Gloudemans,M.J., Fresard,L., Granja,J.M., Louie,B.H., Eulalio,T., Shams,S., Bagdatli,S.T. *et al.* (2020) Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for alzheimer's and parkinson's diseases. *Nat. Genet.*, **52**, 1158–1168.
40. Fang,H., Knezevic,B., Burnham,K.L. and Knight,J.C. (2016) XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med*, **8**, 129.
41. 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.
42. Hu,Y., Li,M., Lu,Q., Weng,H., Wang,J., Zekavat,S.M., Yu,Z., Li,B., Gu,J., Muchnik,S. *et al.* (2019) A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat. Genet.*, **51**, 568–576.
43. Zhou,D., Jiang,Y., Zhong,X., Cox,N.J., Liu,C. and Gamazon,E.R. (2020) A unified framework for joint-tissue transcriptome-wide association and mendelian randomization analysis. *Nat. Genet.*, **52**, 1239–1246.
44. 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.
45. 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.
46. van Rheenen,W., van der Spek,R.A.A., Bakker,M.K., van Vugt,J., Hop,P.J., Zwamborn,R.A.J., de Klein,N., Westra,H.J., Bakker,O.B., Deelen,P. *et al.* (2021) Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat. Genet.*, **53**, 1636–1648.
47. Keller,M.F., Ferrucci,L., Singleton,A.B., Tienari,P.J., Laaksovirta,H., Restagno,G., Chio,A., Traynor,B.J. and Nalls,M.A. (2014) Genome-wide analysis of the heritability of amyotrophic lateral sclerosis. *JAMA Neurol.*, **71**, 1123–1134.
48. Nubling,G.S., Mie,E., Bauer,R.M., Hensler,M., Lorenzl,S., Hapfelmeier,A., Irwin,D.E., Borasio,G.D. and Winkler,A.S. (2014) Increased prevalence of bladder and intestinal dysfunction in amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Frontotemporal Degener*, **15**, 174–179.
49. Lopes de Carvalho,M.L., Motta,R., Battaglia,M.A. and Brichetto,G. (2011) Urinary disorders in amyotrophic lateral sclerosis subjects. *Amyotroph. Lateral Scler.*, **12**, 352–355.
50. Pan,S., Liu,X., Liu,T., Zhao,Z., Dai,Y., Wang,Y.Y., Jia,P. and Liu,F. (2022) Causal inference of genetic variants and genes in amyotrophic lateral sclerosis. *Front. Genet.*, **13**, 917142.
51. Megat,S., Mora,N., Sanogo,J., Catanese,A., Alami,N.O., Freischmidt,A., Mingaj,X., de Calbiac,H., Muratet,F., Dirrig-Grosch,S. *et al.* (2021) Loss of nucleoporin nup50 is a risk factor for amyotrophic lateral sclerosis. medRxiv doi: https://doi.org/10.1101/2021.08.23.21262299, 25 August 2021, preprint: not peer reviewed.
52. Bandres-Ciga,S., Noyce,A.J., Hemani,G., Nicolas,A., Calvo,A., Mora,G., Consortium,I., International,A.L.S.G.C., Tienari,P.J., Stone,D.J. *et al.* (2019) Shared polygenic risk and causal inferences in amyotrophic lateral sclerosis. *Ann. Neurol.*, **85**, 470–481.
53. Peng,B., Yang,Q., R,B.J., Liu,Y., Akbar,M., Song,B.J., Zhou,S. and Wang,X. (2020) Role of alcohol drinking in alzheimer's disease, parkinson's disease, and amyotrophic lateral sclerosis. *Int. J. Mol. Sci.*, **21**, 2316.