



REVIEW

Brain Banks Spur New Frontiers in Neuropsychiatric Research and Strategies for Analysis and Validation



Le Wang^{1,2,#,a}, Yan Xia^{1,3,#,b}, Yu Chen^{1,c}, Rujia Dai^{1,3,d}, Wenying Qiu^{4,e},
Qingtuan Meng^{1,5,f}, Liz Kuney^{3,g}, Chao Chen^{1,6,*,h}

¹ Center for Medical Genetics & Human Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha 410078, China

² Child Health Institute of New Jersey, Department of Neuroscience, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ 08901, USA

³ Psychiatry Department, SUNY Upstate Medical University, Syracuse, NY 13210, USA

⁴ Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100101, China

⁵ Affiliated Hospital of Guilin Medical University, Guilin 541000, China

⁶ National Clinical Research Centre for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha 410000, China

Received 11 January 2019; revised 13 February 2019; accepted 1 March 2019

Available online 5 December 2019

Handled by Hongxing Lei

KEYWORDS

Neuropsychiatric disorders;
Brain bank;
Postmortem brain;
Expression quantitative trait loci;
GWAS interpretation

Abstract Neuropsychiatric disorders affect hundreds of millions of patients and families worldwide. To decode the molecular framework of these diseases, many studies use human **postmortem brain** samples. These studies reveal brain-specific genetic and epigenetic patterns via high-throughput sequencing technologies. Identifying best practices for the collection of postmortem brain samples, analyzing such large amounts of sequencing data, and interpreting these results are critical to advance neuropsychiatry. We provide an overview of human brain banks worldwide, including progress in China, highlighting some well-known projects using human postmortem brain

* Corresponding author.

E-mail: chenchao@sklmg.edu.cn (Chen C).

Equal contribution.

^a ORCID: 0000-0002-6825-6614.

^b ORCID: 0000-0001-6432-9911.

^c ORCID: 0000-0003-2822-0294.

^d ORCID: 0000-0002-9979-3295.

^e ORCID: 0000-0003-4906-8746.

^f ORCID: 0000-0002-6421-162X.

^g ORCID: 0000-0001-9324-5434.

^h ORCID: 0000-0002-5114-2282.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2019.02.002>

1672-0229 © 2019 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

samples to understand molecular regulation in both normal brains and those with neuropsychiatric disorders. Finally, we discuss future research strategies, as well as state-of-the-art statistical and experimental methods that are drawn upon **brain bank** resources to improve our understanding of the agents of neuropsychiatric disorders.

Introduction

Neuropsychiatric and neurological disorders, such as schizophrenia (SCZ), bipolar disorder (BIP), major depression (MD), and Alzheimer's disease (AD), are the leading cause of disability worldwide [1]. However, for more than half a century, a stagnant understanding of their pathophysiology has blocked the development of effective and well-validated neuropsychiatric therapies. Yet, the characteristically high heritability of these disorders should inform us that an earnest understanding of the genetic mechanisms behind these diseases is essential [2,3]. Genome-wide association studies (GWAS) are achieving huge successes in identifying disease-associated variants. For example, the Psychiatric Genomics Consortium (PGC; <http://www.med.unc.edu/pgc>) has identified hundreds of loci associated with SCZ [4], as well as dozens of loci associated with BIP [5] and MD [6,7].

Although many disease-associated variants have been identified, most have small effect sizes and are located in non-coding regions, which hinders interpretation of their functions and disease implications. Quantitative trait loci (QTL) analysis integrates population-based human variation with genome-wide molecular information, such as gene expression [8], DNA methylation [9], histone modifications [10], or chromatin states [11]. QTL is a possible solution for deciphering the function of non-coding variants [12]. Interestingly, most QTL signals show strong tissue specificity [13]. For example, the non-coding variant *rs199347*, associated with Parkinson's disease exclusively, affects the expression of protein-coding gene *GPNMB* (*Glycoprotein Nmb*) in the human brain while sparing other tissues [14]. Robust brain bank collections can facilitate the comprehensive molecular profiling needed to advance research in neuropsychiatric disorders.

Many prominent brain projects on neuropsychiatric disorders generated big data at multiple regulatory levels, including epigenetic markers and gene expression. Although these multidimensional data identified numerous functional genomic elements, challenges remain that impede our full understanding of the underlying molecular etiologies of neuropsychiatric disorders and limit our ability to translate this understanding into improving human health. Although brain tissue samples have become a critically valuable resource for neuropsychiatric studies, to our knowledge, there are only a few comprehensive reports on brain bank resources. Therefore, in this review, we present a summary of the most representative brain banks and brain projects, emphasizing how harnessing these new resources and technologies can refine our insight into the underlying mechanisms of neuropsychiatric disorders. For example, we will discuss brain expression quantitative trait loci (eQTL) analysis as a methodology to interpret the potential functions of GWAS signals identified in various brain disorders. We also discuss the insights and limitations of current brain studies. Finally, we propose best practices for analyzing postmortem brain samples to more accurately interpret the

resulting multidimensional data, thereby augmenting future investigations.

Brain banks

A brain bank is a centralized resource that collects and stores postmortem brain tissues. Brain banks share samples and clinical information with qualified researchers worldwide to advance brain studies in both basic research and clinical trials. Currently, hundreds of human brain banks worldwide are dedicated to the collection of human post-autopsy brain tissues [15]. These have been helpful in demystifying brain-related diseases, such as AD, SCZ, BIP, and MD. Although brain tissue collection is the cornerstone for brain studies, obtaining high-quality brain tissues can be problematic. To counter this and enable better access, large networks such as the Australian Brain Bank Network, BrainNet Europe [16], NeuroBioBank [17], and the UK Brain Banks Network, share technologies and brain sample information. These brain banks have collectively standardized disease diagnosis and tissue collection procedures [18]. Here, we introduce procedures for obtaining high-quality postmortem brain tissue followed by a brief overview of brain banks worldwide and in China.

Working with high-quality postmortem brain tissues

Various factors critically impact the quality of postmortem brain samples [19]. For example, an extended time interval between death and acquisition, the postmortem interval (PMI), can lead to RNA degradation [20]. Effective and rapid brain tissue acquisition and long-term preservation requires precise and unified manipulation using anatomical, cryopreservation, and slicing technologies. Rapid autopsy programs based on round-the-clock autopsy greatly shorten the PMI. Many important parameters are used to determine brain tissue quality, including brain pH, as well as the integrity of DNA, RNA, and proteins [19]. In a strict autopsy environment, which often prolongs the process of sample acquisition, brain pH can notably affect the integrity of RNA and DNA [19]. While formalin-fixed samples tender brain DNA relatively efficiently, the yields of high-quality RNA is somewhat problematic. It is clear that acquiring and preserving high-quality postmortem brain tissues requires great skill and adherence to standard procedures.

Accurately segmenting brain regions is critical, since biological functions vary by brain regions. There are several brain regions highly related to neuropsychiatric cognitive and emotional dysfunction. For example, the dorsolateral prefrontal cortex (DLPFC) and the hippocampus manage cognitive processes including working memory, planning, and cognitive flexibility. The striatum can receive glutamatergic and dopaminergic inputs from multiple sources functional, in the cognitive and reward systems. Accurate definitions for landmarks and label boundaries are important based on our

assumption of the close correspondence of brain function to anatomy. The human cerebral cortex is difficult to label due to the great anatomical variations in the cortical folds and the difficulties in establishing consistent and accurate reference landmarks across the brain. Brain banks classify brain regions according to the Brodmann atlas, which defines 52 cerebral cortex regions [21]. Although there are no clear ‘gold standards’ for measuring the accuracy of anatomical assignments, it is common to measure consistency across trained human observers and variability across co-registered landmarks.

Brain banks worldwide

Although the study of human brains is as old as medicine, brain banks benefitting neuropsychiatric research today arise from international collaboration, guided by modern principles of ethics, quality, and safety with valid scientific aims. One of the most famous brain banks is the Netherlands Brain Bank (NBB) in Amsterdam (<https://www.brainbank.nl/>) [16]. The NBB was established in 1985 to collect human brain tissues from donors with various neurological and psychiatric disorders and also non-diseased donors. NBB had collected brain samples from more than 4000 donors. Launched in 2001, the BrainNet Europe consortium (<https://www.neuropathologie.med.uni-muenchen.de/funktionen/bne/index.html>) has 19 members from across the continent. The brain tissues and the corresponding anonymized summary of each donor’s medical records support extensive national and international research projects. North America with a wealth of brain banking resources has over 50 brain banks including the Allen Institute for Brain Science (<https://alleninstitute.org/>), Harvard Brain Tissue Resource Center (<https://hbtrc.mclean.harvard.edu/>), and the Stanley Medical Research Institute (<http://www.stanleyresearch.org/>). Representative brain banks also include the New South Wales Tissue Resource Centre (Australia, <https://nswbrainbank.org.au/about/nswbtrc>), Tokyo Metropolitan Institute of Gerontology (Japan, <http://www.tmig.or.jp/>), and the Brain Bank of the Brazilian Aging Brain Study (Brazil, http://www2.fm.usp.br/gerolab_en/index.php).

Brain banks in China

In China, the number of brain samples is quite limited. The creation of Chinese brain banks has recently become a priority for researchers. China’s Han population represents the world’s largest ethnicity and roughly 80% of East Asia’s population; yet brain data from this population is currently understudied and will prove a valuable resource within the global survey. However, brain banking in China is slowly developing, with the China Human Brain Banking Consortium established in 2014 at the International Workshop on Human Brain Banking in China [22]. So far, there are nearly one thousand brain samples from dozens of consortium members, including the Xiangya School of Medicine Brain Bank, the Zhejiang University of China Brain Bank, the Chinese Academy of Medical Sciences & Peking Union Medical College Human Brain Bank, and others. The consortium organizes conferences and workshops annually to build up a unified process for brain tissue acquisition and storage, discussing policy for sample sharing, and exchanging experiences and new findings [23].

Evolutionary perspectives can help us better understand the relationship between brain development and disease. Therefore, nonhuman primate (NHP) brain resources play an important role in distinguishing human brain-specific regions. The Nonhuman Primate Reference Transcriptome Resource (<http://nhprtr.org/index.html>) began in 2010 [24]. Its goal is to establish an NHP reference transcriptome consisting of transcriptome sequencing data from multiple nonhuman species, including *Papio anubis*, *Pan troglodytes*, *Macaca fascicularis*, *Gorilla gorilla*, and 11 other non-human primates. Within their protocol, 22 tissue types are collected from four brain regions (*i.e.*, cerebellum, frontal cortex, hippocampus, and temporal lobe). By comparing brain regions of humans to those of non-human primates, Doan *et al.* was able to identify human-specific social and behavioral traits associated with autistic spectrum disorder (ASD) that are regulated by the human accelerated genomic regions [25].

Brain projects

The collective increase in brain banks globally has spurred a multitude of brain research projects. For most projects, samples are obtained from well-constructed brain banks [26]. Brain research projects focus on many different dimensions, including brain development, spatiotemporal gene expression, epigenetic modification, and pathological characterization of neuropsychiatric disorders. Some of these efforts include, BrainSpan (<http://www.brainspan.org/>) [27,28], UK Brain Expression Consortium (UKBEC, www.braineac.org/) [29], Genotype Tissue Expression Project (GTEx, <https://gtexportal.org/>) [30], CommonMind Consortium (CMC, commonmind.org/) [31], BrainSeq (<http://eqtl.brainseq.org/>) [32], the Religious Orders Study and Memory and Aging Project (ROSMAP, <http://www.radc.rush.edu/>) [33], PsychENCODE (<http://psychencode.org/>) [34], and BrainCloud (<http://braincloud.jhmi.edu/>) [35]. They aim to gather genotypic data and data at other regulatory levels for the human brain, to reveal the genetic regulatory mechanisms of the human brain at different levels (Figure 1 and Tables 1–3).

Benefitting from the continual production of data and strengthened by in-depth structured analyses, brain projects are valuable references revealing basic functions as well as molecular and cellular pathologies related to neuropsychiatric disorders. As a source of data, each brain project offers unique design features and advantages for specific research aims. For instance, the GTEx project, which collects samples from non-disease tissue sites, including but not limited to the brain, focuses on tissue specificity of gene expression, cross-tissue gene expression regulation, and genetic variations that contribute to complex diseases and quantitative traits in humans [30]. The UKBEC, which collects samples from across a wide-range of brain regions, up to 12 regions per donor, focuses on the regulation and alternative splicing of gene expression [29]. BrainCloud [35] and BrainSpan [27,28] focus on spatiotemporal gene expression regulation during the development of the human brain from embryonic to adult stages. Although BrainCloud is superior in terms of sample size, BrainSpan includes more brain regions and types of sequencing data, such as miRNA expression.

Other brain projects include samples from donors with or without neuropsychiatric disorders, exploring the differences

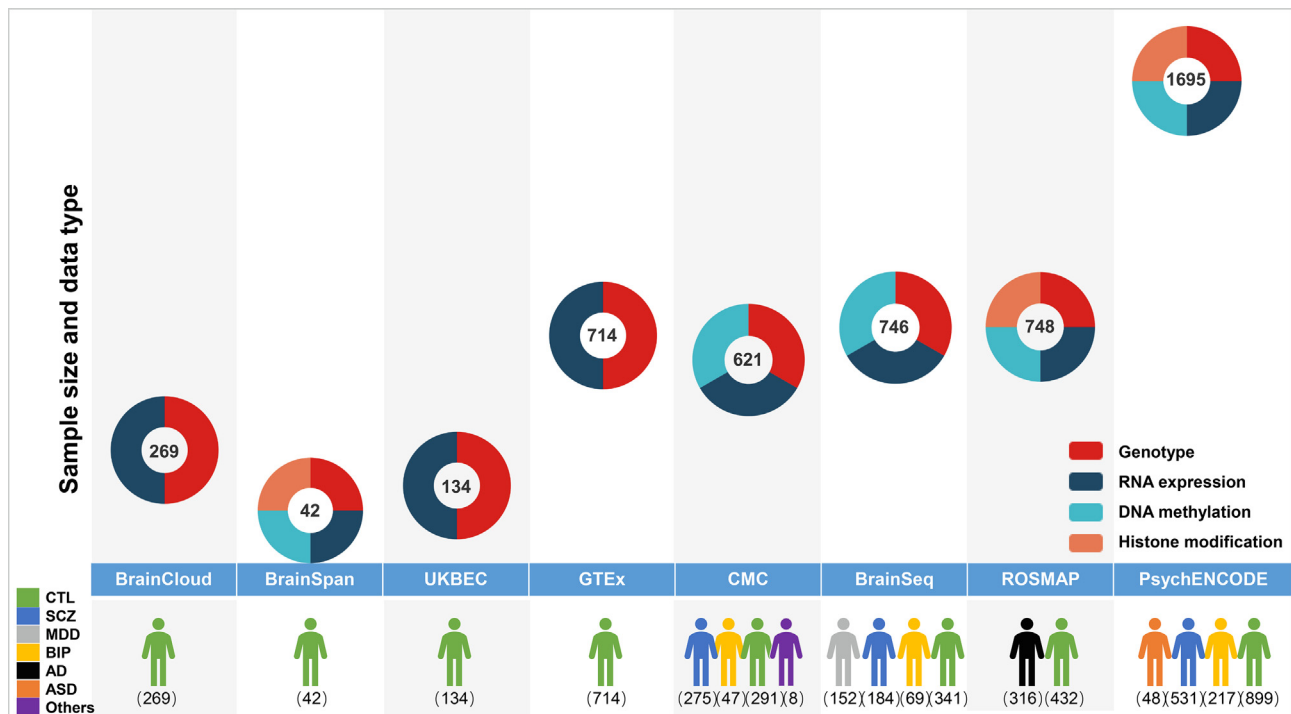


Figure 1 Overview of the representative brain projects

Numbers in cycles indicate the number of brain samples used in each project. Different data types are indicated using different colors, which include genotype, RNA expression, DNA methylation, and histone modification data. Colors in the bottom panel indicate the distribution of healthy controls or patients with different diseases included in the respective projects. The projects and their web links for access were listed below. BrainCloud (<http://braincloud.jhmi.edu/>) [35]; BrainSpan (<http://www.brainspan.org/>) [27,28]; UKBEC, UK Brain Expression Consortium (www.braineac.org/) [29]; GTEx, Genotype Tissue Expression Project (<https://gtexportal.org/>) [30]; CMC, CommonMind Consortium (commonmind.org/) [31]; BrainSeq (<http://eqtl.brainseq.org/>) [32]; ROSMAP, the Religious Orders Study and Memory and Aging Project (<http://www.radc.rush.edu/>) [33]. Only Capstone 1 data from PsychENCODE (<http://www.psychencode.org/>) were summarized in this figure. PsychENCODE Capstone 1 data comprise BrainGVEX, BrainSpan, CommonMind, UCLA- ASD, Yale- ASD, BipSeq, LIBD szControl, and CMC_HBCC datasets, but does not include fetal brain samples and outliers. CTL, control; SCZ, schizophrenia; MDD, major depressive disorder; BIP, bipolar disorder; AD, Alzheimer's disease; ASD, autism spectrum disorder.

Table 1 Number of individuals across developmental stages per brain project

Stage	Age	BrainCloud	BrainSpan	UKBEC	GTEx	CMC	BrainSeq	ROSMAP	PsychENCODE
Fetal	~0	38	19	0	0	0	56	0	0
Infancy and childhood	0–12	34	12	0	0	0	31	0	65
Adolescence	12–20	49	4	2	0	3	60	0	88
Young adulthood	20–40	53	5	18	107	60	179	0	302
Middle adulthood	40–60	73	2	49	357	163	320	0	606
Late adulthood	≥ 60	22	0	65	250	395	100	748	634

Note: Only Capstone 1 data from PsychENCODE were summarized in this table. PsychENCODE Capstone 1 data comprise BrainGVEX, BrainSpan, CommonMind, UCLA- ASD, Yale- ASD, BipSeq, LIBD szControl, and CMC_HBCC datasets, but does not include fetal brain samples and outliers. UKBEC, UK Brain Expression Consortium; GTEx, Genotype Tissue Expression Project; CMC, CommonMind Consortium; ROSMAP, the Religious Orders Study and Memory and Aging Project.

Table 2 Number of individuals by race per brain project

Race	BrainCloud	BrainSpan	UKBEC	GTEx	CMC	BrainSeq	ROSMAP	PsychENCODE
European	112	21	134	608	500	–	730	1272
African American	147	14	0	91	90	–	14	350
Hispanic	6	4	0	0	26	–	0	41
Asian	4	1	0	8	4	–	3	20
Others	0	2	0	7	1	–	1	5

Note: “–”, data not available.

Table 3 Number of samples per brain region per brain project

Brain region	BrainCloud	BrainSpan	UKBEC	GTEX	CMC	BrainSeq	ROSMAP	PsychENCODE
Prefrontal cortex	269	37	127	129	621	746	748	1695
Temporal cortex	0	39	119	0	0	0	0	134
Anterior cingulate cortex	0	37	0	121	0	0	0	0
Cerebellum	0	35	130	173	0	0	0	0
Hippocampus	0	37	122	0	0	270	0	0
Caudate	0	0	0	160	0	500	0	0
Amygdala	0	36	0	100	0	0	0	0
Hypothalamus	0	0	0	121	0	0	0	0
Nucleus accumbens	0	0	0	147	0	0	0	0
Putamen	0	0	129	124	0	0	0	0
Substantia nigra	0	0	101	88	0	0	0	0

Note: Samples from BrainSpan, UKBEC, GTEX, BrainSeq, and PsychENCODE datasets were collected from multiple brain regions per individual.

between brain features of patients and those of controls. The Religious Orders Study (ROS) [36] and the Memory and Aging Project (MAP) comprise the ROSMAP project [37], a longitudinal, clinical, and pathological cohort study of aging and dementia. The ROS component focuses on data from various conditions of dementia within a limited population, while the MAP project focuses on reduced cognitive and motor function and disease risk of those with AD within a more varied population. CMC and BrainSeq [31,32] focus on neuropsychiatric disorders, including SCZ, BIP, ASD, and MD, by comparing diseased samples with controls. The BrainSeq project seeks to identify therapeutic drug targets for neuropsychiatric disorders by understanding the genetic and epigenetic regulations across the human lifespan. The PsychENCODE project [34] makes an extensive, "multidimensional" genetic and epigenetic dataset available to the public, derived from the tissue samples of postmortem healthy and diseased human brains. The project characterizes disease-associated regulatory and genetic features within pathological models, focusing initially on ASD, BIP, and SCZ [38–40]. Current data generated from the PsychENCODE project include: chromatin immunoprecipitation following next-generation sequencing (ChIP-seq), RNA-seq, whole-genome bisulfite sequencing (WGBS), miRNA sequencing (miRNA-seq), isoform sequencing (IsoSeq), assay for transposase accessible chromatin with high-throughput sequencing (ATAC-seq), enhanced reduced representation bisulfite sequencing (ERRBS), single nucleotide polymorphism (SNP) genotypes, array methylation, and reverse phase protein array (RPPA).

The major findings using postmortem samples from brain projects are summarized in Table S1. These data provide important insights into the contribution of genetic and epigenetic factors to mechanisms underlying neuropsychiatric disorders. Particularly, the BrainSeq Consortium performed RNA-seq on 495 postmortem brains with ages across the human lifespan, including 175 samples from SCZ patients and 320 controls [41]. Through integrative analyses, this consortium demonstrates that 48.1% SCZ GWAS risk variants are associated with expression of nearby genes, and 237 differentially expressed genes implicated in synaptic processes are regulated in early brain development. The earlier study on the epigenetic landscape of frontal cortex in patients with SCZ [42] shows that SCZ-associated CpGs strongly correlate with fetal development stage rather than the adult stage of the brain. These results reveal potential SCZ pathogenesis in gene expression

and DNA methylation during brain development and maturation. Moreover, recent studies by the PsychENCODE project have identified cell composition and maturation leading to spatiotemporal transcriptomic variation patterns in human and macaque brains [43]. They also observe associations of neuropsychiatric diseases with epigenetic markers [38], QTLs [39], and isoform-level changes [44]. For example, they have identified several interesting targets, including *DGCR5* and *POU3F2*, which play essential roles in regulating SCZ-related genes at the network level [45,46]. These postmortem studies provide important insights into the genetic architecture for robust and informative models of neuropsychiatric disorders, which will help in devising strategies for novel therapeutic interventions.

Strategies and execution

Unarguably, postmortem brain resources are valuable in revealing the biological underpinnings of neuropsychiatric disorders; however, unravelling the full potential of multidimensional brain data is still a great challenge. One promising strategy employs QTL analysis, which integrates population-based human variations with genome-wide molecular information (e.g., gene expression, DNA methylation, histone modification, and chromatin states). Widely used, QTL captures the associations between genetic variants and gene expression. For instance, QTL can be used to investigate variants at *cis*-regulatory elements, such as transcription factor-binding regions, which confer differential expression of target genes. Combined with GWAS, QTL studies interpret how disease-associated variants may contribute to molecular traits and disease susceptibility. In this section, we will discuss eQTL specifically, summarizing the key steps for pre-processing of brain gene expression data, highlighting important issues in eQTL analysis, explaining how to use eQTL to interpret GWAS signals, and finally, introducing cutting-edge experiments to validate regulatory signals (Figure 2 Overflow of the research strategies and methods).

Pre-processing brain gene expression data

Although laborious, data pre-processing is essentially the first step to ensure proper and efficient data modelling. A clean, software-compatible format will ensure reproducible results

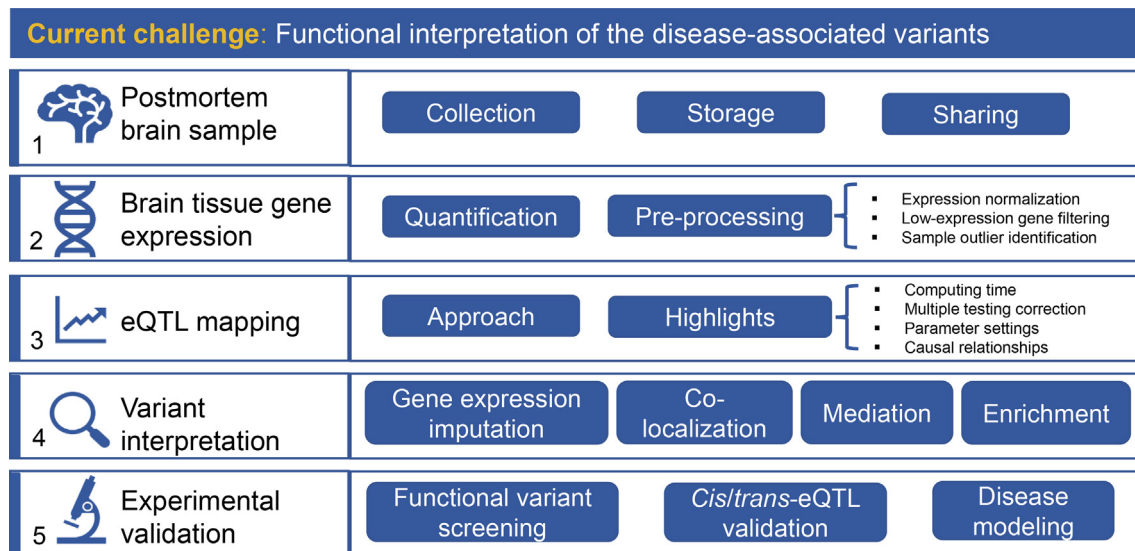


Figure 2 Overview of strategies and methods in neuropsychiatric studies

and save hours, even days, of data analysis [47]. Variable reporting of gene expression can arise from biological factors and technical variations. To distinguish biological variations from confounding factors, technical factors (*e.g.*, batch effects) must be removed or adjusted. Major pre-processing steps include gene expression normalization and filtering, sample outlier identification, and covariate correction. Because strategies in the human brain studies are the major focus of this article, we will only cover the key steps that may alter the quality of brain gene expression results. Comprehensive guidelines for gene expression data analysis are well discussed elsewhere [48,49] and are beyond the scope of this review.

The first key step is gene quantification and filtering. Tools for quantification are widely available, such as Cufflinks [50], eXpress [51], Flux Capacitor [52], kallisto [53], RSEM [54], Sailfish [55], and Salmon [56]. Each tool can accurately assign reads to transcripts and quantify expression. These functions are vital for interpreting tissue-specific expression patterns in the brain [57]. However, the criteria for poorly expressed genes vary across studies. For instance, PsychENCODE project filters genes with transcript per million (TPM) < 0.1 in more than 25% of samples [58].

The second key step is sample outlier removal. Samples with a high degree of poorly expressed genes or gene expression patterns distinct from other samples are removed. This step can be carried out in dimension reduction analysis such as principal component analysis (PCA) and multidimensional scaling (MDS). Network concepts such as standardized connectivity (the overall strength of connections between a given sample and all of the other samples in a network) are also used to confirm sample outliers within a group [59].

The third key step is controlling covariates, including both known and unknown covariates. Known covariates can be either technical, such as batch effects, or biological, such as sex and age. Some biological covariates have been ignored by earlier research, leading to potentially confounding results. For instance, cell-type composition is one such common problem: since bulk-tissue RNA-seq only measures the average behavior, it is unable to capture cellular heterogeneity, which

makes the observed changes in gene expression reflect only changes in cell-type composition, rather than fundamental changes in cell states [60]. Therefore, cell numbers and ratios of multiple cell types are important biological covariates, that affect brain gene expression profiles, since different cell states rather than cell type composition reflect distinct biological activities and gene expression patterns. Another covariate that is critical but often neglected is drug treatment history. Gene expression can vary dramatically across therapeutic courses. The unknown factors, also called hidden determinants, can reduce the power to find eQTLs. Surrogate variable analysis (SVA) [34] or probabilistic estimation of expression residuals (PEER) [61] can calculate unknown sources of variation, followed by a linear regression model to remove them. One could choose ComBat [62] (in R package *sva*) to remove the batch effects; finally, a linear regression model will remove the confounding factors.

Pitfalls and promises in eQTL analysis

The aim of eQTL analysis or eQTL mapping is to characterize associations between the expression of corresponding genes and SNPs, thereby isolating specific regulatory regions within the genome. A variety of approaches have been proposed, including using linear regression, ANOVA, and non-linear models. Some approaches also account for pedigree and other confounding factors [63], integrating known functional elements [64], or considering allelic imbalances [65]. FastQTL, for instance, features expansive permutations that refine *P* values and reduce computational burden.

Several issues should be highlighted in eQTL analysis. The first is computing time. Pairwise association compares up to one million genetic variants to tens of thousands of genes, making analysis computationally intensive, especially when employing a non-linear model on a larger dataset. Secondly, multiple testing corrections become necessary for many of the tests performed. One common solution is to calculate the false discovery rate for each SNP-gene pair. Furthermore, separating the *cis*-eQTLs and *trans*-eQTLs is crucial, since local

variants may regulate gene expression much more than distal variants. However, this correction alone is too strict because those tests are not biologically independent. Therefore, permutation-based methods, which create the null distribution of associations by tens of thousands of permutations, were developed to set up an effective threshold for identifying statistically significant eQTLs. Third, parameter settings can be a critical factor when comparing eQTLs across multiple studies. For example, the distance between SNPs and gene locations is used to differentiate *cis*-eQTL and *trans*-eQTL signals, which could be defined as 1 Mb, 5 Mb or 10 Mb in different studies. Varied distance settings may lead to different statistical burdens for SNPs located in regions ranging from 1 to 10 Mb and result in variable outcomes. The customized cut-off threshold for minor allele frequency (MAF) may also cause the loss of some true signals. Fourth, some eQTLs have such strong correlations with gene expression that they may not prompt gene expression changes. In other words, those genetic variants may be correlated with the causal variants due to linkage disequilibrium or other factors. Both statistical and experimental approaches have been proposed to solve this problem [66,67]; either ways, it is critical to identify true causal variants when integrating eQTL and GWAS results [68].

Interpreting GWAS signals

GWAS variants can increase or decrease gene expression, a culprit behind the etiology of many diseases; QTL helps us interpret how non-coding GWAS variants work. Several kinds of methods, each with unique principles, have been developed to integrate GWASs and eQTL results (Table 4). One type of method is based on gene expression imputation, such as PrediXcan [70] and transcriptome-wide association study (TWAS/FUSION) [71]. These methods estimate the genetically regulated component of expression using reference transcriptome datasets such as GTEx [30], GEUVADIS [8], and DGN [85] among others to build a database of prediction models. For each new genotype data, these methods impute gene expression and then correlate that gene expression to a trait of interest to identify trait-associated genes. The second group investigates the co-localization of GWAS causal variants and eQTL causal variants. For example, COLOC [72], MOLOC [73], ENLOC [86], HyPrColoc [74], and Sherlock [75] use a Bayesian statistical framework to integrate GWAS summary data and eQTLs to estimate the causal variants, and eCAVIAR [78] considers multiple causal variants within one locus. Other groups include enrichment methods, such as S-LDSC [82] and eQTLEnrich [81], and mediation methods. Summary data-based Mendelian Randomization (SMR) [66] and generalized SMR (GSMR) [84] test whether the effect of a GWAS SNP on a specific trait has been mediated by the expression of a gene.

While using eQTL to interpret GWAS results is a good way to understand gene regulatory mechanisms, it is not without limitations. First, for some diseases if the most relevant tissue/cell types or developmental stages are not available in eQTL analysis, we can find neither the true genetic regulation nor the related genes. Second, gene expression is only one dimension of genetic regulation. If the biological mechanism is independent of gene expression levels but affects other regulatory cascades, such as splicing, chromosome accessibility, or ribo-

some profiling, eQTL alone will not be enough to explain the underlying processes. Third, QTL and GWAS focus on common variants, therefore they cannot capture rare variants with higher effect sizes in gene expression [87].

Experimental approaches to characterize functional variants

After identifying disease risk variants or regulatory elements using the aforementioned bioinformatics analysis methods, the next step is to characterize the function of the variants.

To validate risk variants as the eQTL signal, using high-throughput and sensitive methods to measure their effect on gene expression is a widely adopted approach. As a favored method, reporter gene assay screening validates whether functional elements with eQTL signals regulate target gene expression, by cloning the regulatory elements into an expression reporter vector [74]. Whereas reporter assays validate regulatory functions of variance *in vitro*, CRISPR can be used to validate regulatory functions of the variance within native chromosome regions *in vivo*. For instance, Diao et al. used a CRISPR tiling-deletion-base genetic approach to identify some *cis*-regulatory elements in mammalian cells [88]. Furthermore, high throughput CRISPR screening systems, such as the CRISPR-Cas9, have been used to investigate the effect of the regulatory variance on the downstream target genes [75,78,81,82,84]. Recently, studies have refined the resolution of this technique, including the dCas9 fusion APOBEC1 (Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 1)/TadA (tRNA-specific adenosine deaminase)-mediated efficient single base mutation system [69,87]. While CRISPR technology has these advanced capabilities, it is not without limitations. For instance, inconsistencies such as off-target genome editing (*i.e.*, inducing unwanted allelic variances) have been problematic to date [89]. Nonetheless, CRISPR has tremendous potential for single base screening and clinical applications. We are confident that CRISPR will mature into a dependable tool for correcting genetic variation in the future.

To understand the influence of risk variants on gene expression, several productive tools have been developed. For the chromatin states, ChIP-seq is an efficient genome-wide method to identify the transcription factor binding sites in open chromatin regions, including promoter, enhancer and other transcription active elements. Based on the principle of ChIP-seq, a series of targeted chromatin DNA sequencing technologies have been developed (*e.g.*, DNase-seq, MNase-seq, FAIRE-seq and ATAC-seq). For example, Forrest et al. revealed the function of non-coding GWAS risk variants using ATAC-seq data from neurons derived from SCZ patient induced pluripotent stem cells (iPSCs) [90]. Chip-related technology can help us to annotate and interpret the functionality of disease-associated non-coding variants. Data on DNA-protein binding generated by sequencing technologies requires validation using *in vitro* methods, including the electrophoretic mobility shift assays (EMSA). However, the throughput of the EMSA-based experiments is limited. To improve the throughput of this *in vitro* validation, mass spectroscopy proteome-wide analysis of SNPs (PWAS) can be applied for screening genetic variants for differential transcription factor binding [91].

Risk variants located in the untranslated region (UTR) and intronic regions may also contribute to disease through post-transcriptional regulation, such as splicing, RNA stability, or

Table 4 Algorithms and software for integrating GWAS and eQTL data

Name	Description	LD considered	Programming language	Operating system	Link	Ref.
MetaXcan	Gene expression imputation	Y	Python	Unix/Linux	https://github.com/hakyimlab/MetaXcan	[69]
PrediXcan	Gene expression imputation	N	Python	Unix/Linux	https://github.com/hakyimlab/PrediXcan	[70]
TWAS / FUSION	Gene expression imputation	Y	R	Unix/Linux, Mac OS, Windows	http://gusevlab.org/projects/fusion/	[71]
COLOC	Co-localization	Y	R	Unix/Linux, Mac OS, Windows	https://github.com/chriswallace/coloc	[72]
MOLOC	Co-localization	Y	R	Unix/Linux, Mac OS, Windows	https://github.com/clagiamba/moloc	[73]
ENLOC/fastENLOC	Co-localization	Y	Perl	Unix/Linux, Mac OS, Windows	https://github.com/xqwen/integrative	[73]
HyPrColoc	Co-localization	Y	R	Unix/Linux, Mac OS, Windows	https://github.com/jrs95/hyprcoloc	[74]
Sherlock	Co-localization	Y	-	Web interface	http://sherlock.ucsf.edu/	[75]
JEPEG	Joint eQTL analysis	Y	C++	Unix/Linux	https://dleelab.github.io/jepeg/	[76]
CAVIAR	Co-localization	Y	C	Unix/Linux, Mac OS, Windows	http://genetics.cs.ucla.edu/caviar/	[77]
eCAVIAR	Co-localization	Y	C	Unix/Linux, Mac OS, Windows	http://genetics.cs.ucla.edu/caviar/	[78]
GMAC	Mediation analysis	N	R	Unix/Linux, Mac OS, Windows	https://cran.r-project.org/web/packages/GMAC	[79]
FINEMAP	Co-localization	Y	C	Unix/Linux, Mac OS	http://www.christianbenner.com/	[80]
eQTLEnrich	Enrichment	Y	MATLAB	Unix/Linux, Mac OS, Windows	https://segrelab.meei.harvard.edu/software/	[81]
S-LDSC	Enrichment	Y	Python	Unix/Linux, Mac OS, Windows	https://github.com/bulik/ldsc	[82]
NEO	Structural equation model	N	R	Unix/Linux, Mac OS, Windows	https://labs.genetics.ucla.edu/horvath/htdocs/aten/NEO/	[83]
SMR	Mendelian randomization	Y	C	Unix/Linux, Mac OS, Windows	http://cnsgenomics.com/software/smr	[66]
GSMR	Mendelian randomization	Y	R	Unix/Linux, Mac OS, Windows	http://cnsgenomics.com/software/gsmr/	[84]

Note: eQTL, expression quantitative trait loci; TWAS, transcriptome-wide association study; JEPEG, joint effect on phenotype of eQTLs/functional SNPs associated with a gene; CAVIAR, causal variants identification in associated regions; eCAVIAR, eQTL and GWAS causal variants identification in associated regions; GMAC, Genomic Mediation Analysis with Adaptive Confounding Adjustment; NEO, Network Edge Orienting; SMR, summary-data-based Mendelian randomization; GSMR, generalized summary-data-based Mendelian randomization.

non-coding regulation. High-throughput analysis of RNA isolated by cross-linking immunoprecipitation sequencing (CLIP-Seq) could be used to map protein-RNA binding site or RNA modification site *in vivo* [92–94]. This technique can reveal risk variants that affect gene expression at the post-transcriptional level. For example, Eric T. Wang used RNA-seq and CLIP-seq to reveal the transcriptome-wide regulation of pre-mRNA splicing and mRNA localization in myotonic dystrophy [95].

It is important to note that risk variants may not necessarily affect expression of the nearest gene. Disease risk variants may also affect expression of distal genes through long-range chromatin interactions [96–98]. The interaction of chromatin-specific regions can be explored by classic chromatin conformation capture (3C) techniques. This 3C-based technology involves cross-linking chromatin interaction sites, using genome DNA cleavage with a restriction enzyme and a ligation reaction to join cross-linked DNA fragments. Chromatin interactions at specific candidate loci could be further validated by polymerase chain reaction (PCR) [99]. For example, Panos Roussos et al. demonstrated physical interactions between the *CACNA1C* eQTL risk locus and distal regulatory elements using 3C techniques in prefrontal cortex [100].

The next step is to explore disease-associated phenotypes of genetic risk variants by establishing cellular models or animal models. For example, human iPSCs (hiPSCs) research detects molecular and cellular phenotypes (*e.g.*, migration, proliferation, and electrophysiology) together with the genetic background of specific patients. Moreover, the 3D culturing of pluripotent stem cells produces organoids, demonstrating their remarkable capacity for self-organization and differentiation. This approach can be used to study human brain specific features and the mechanism of neurodevelopment and neuropsychiatric disorders. For example, Marina Bershteyn et al. used human-derived cerebral organoids to model the cellular features of Miller-Dieker syndrome caused by 17p13.3 deletion [101]. While animal models differ from humans in terms of genetic background, they resemble the spectrum of human disease phenotypes, ranging from tissue and organ to behavior. Those two models, when combined with postmortem brain data, may unlock the mysteries of risk variant function and increase the probability of decoding the pathology of neuropsychiatric diseases.

Future directions

In this review, we summarized the most representative brain banks and brain projects worldwide, supporting a multidimensional understanding of neuropsychiatric disorders from pathology, genetic, and gene expression perspectives. Brain banks and projects are establishing research resources and building coalitions to reduce the incidence and impact of neuropsychiatric disorders. Multidimensional data collected using brain bank resources facilitate the study of complex neuropsychiatric disorders, as brain banks are increasingly linked to important sources of clinical information. Different brain projects use brain bank samples to generate a wide spectrum of data types and serve as an important resource to promoting brain research. Developing advanced research methods and experimental validation of findings increases our capability of finding true causal signals of neuropsychiatric illnesses.

Postmortem brain samples have lent profound insight into genomic, transcriptomic and epigenomic studies, however brain disorder research faces many challenges. Various cell types from different brain regions form specific neural circuits that govern complex behaviors. Most brain studies include samples from different brain regions and use the bulk brain tissue as a whole, which obviously contains many cell types, such as neurons, astrocytes, microglia, and oligodendrocytes. Single-cell studies are increasingly needed to achieve higher resolution in detailed genomic insights. Some recent studies have been used single-cell methods to isolate specific cell types from healthy human brain tissue to characterize human brain development [102,103]. Heterogeneity in medical treatment is one confounding factor that can affect gene expression profiles and some epigenetic marks. Almost every psychotic patient has a long history of drug therapy, but individuals without neuropsychiatric disorders may not, which may result in possible false-positive findings. Furthermore, integrating the drug history relies on obtaining hospital medical records or self-reporting, both of which can be unreliable. For example, patients may refuse to take prescribed medications, while others may not be able to accurately recall their medication history. Directed toxicology testing for each sample is the best solution but may not be practical due to the many types of antipsychotic drugs available and the high expense involved. Moreover, smoking and drinking history, state of death (*e.g.*, unexpected death, expired while asleep, unconsciousness, fever and hypoxia) are also confounding factors for postmortem gene expression and other studies [104,105]. Consider this necessary information when collecting samples.

One vital but challenging aspect of brain collection is the use of fetal and infant brains. In most banks, donated brains come from aged individuals, appropriate for the research of neurodegenerative diseases. For neurodevelopmental diseases, such as autism, SCZ, and intellectual disability, however, fetal and infant brain samples are critical for investigating disease etiology. So far, only a few banks have prenatal samples, and their samples sizes are relatively small. Including fetuses with lethal defects and those with defects not affecting brain function, identified through prenatal genetic screening, could increase available resources. Another solution would be using iPSC-derived neurons or other brain cells to model the very early stages of brain development. Combining these strategies, we can characterize the temporal regulatory landscape of brain development and genomic aberrations related to psychiatric illnesses.

Recently, it has been suggested that all postmortem brain studies are underpowered to correct for genetic and phenotypic heterogeneity [106]. This begs the question, how can these studies derive from the brain banks with limited sample sizes achieve enough statistical power? One solution is in more accurately defining disease-related phenotyping and levels of disease taxonomy. For example, in BIP, only about 30% of patients respond to lithium [107,108], and a portion of patients have DLPFC or hippocampal volume abnormalities [109–112]. Classification of these disease subtypes improves the understanding of disease phenotype. Availability of shared data is another big issue often limiting the power needed for research into neuropsychiatric disorders. With more and more data generated and released, an open public and user-interactive data center is needed to collect and to manage all the repositories. Our group established the Brain EXPression Database

(BrainEXP, <http://www.brainexp.org/>) focusing on brain gene expression patterns in various regions, by sex and age [113]. This database currently includes 4567 brain samples of 2863 normal individuals and will integrate approximately the same number of patient samples in the near future. These combined efforts hold the promise of powering brain studies adequately.

In conclusion, given the expanding framework of brain bank and brain project networks, we can improve exploration into the molecular regulatory mechanisms of neuropsychiatric disorders and facilitate research toward new avenues of treatment.

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31571312, 31970572, and 81401114), the National Key R&D Project of China (Grant No. 2016YFC1306000), Innovation-driven Project of Central South University (Grant Nos. 2015CX5034 and 2018CX033) to CC. We thank Dr. Chunyu Liu from SUNY Upstate Medical University, Dr. Xiaoxin Yan from Central South University, Dr. Chao Ma from the Chinese Academy of Medical Sciences and Peking Union Medical College, and Dr. Zhiping Pang from Child Health Institute of New Jersey for critical reading and comments.

Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2019.02.002>.

References

- [1] Hyman SE. Revolution stalled. *Sci Transl Med* 2012;4:155cm11.
- [2] Cardno AG, Gottesman II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* 2000;97:12–7.
- [3] Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 2003;60:1187–92.
- [4] Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* 2010;68:182–6.
- [5] Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 2019;51:793–803.
- [6] Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018;50:668–81.
- [7] CONVERGE Consortium. Sparse whole genome sequencing identifies two loci for major depressive disorder. *Nature* 2015;523:588.
- [8] Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;501:506–11.
- [9] Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, et al. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet* 2010;86:411–9.
- [10] Waszak SM, Delaneau O, Gschwind AR, Kilpinen H, Raghav SK, Witwicki RM, et al. Population variation and genetic control of modular chromatin architecture in humans. *Cell* 2015;162:1039–50.
- [11] Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* 2015;162:1051–65.
- [12] Li M, Wu DD, Yao YG, Huo YX, Liu JW, Su B, et al. Recent positive selection drives the expansion of a schizophrenia risk nonsynonymous variant at *SLC39A8* in Europeans. *Schizophr Bull* 2016;42:178–90.
- [13] Ongen H, Brown AA, Delaneau O, Panousis NI, Nica AC, Consortium GT, et al. Estimating the causal tissues for complex traits and diseases. *Nat Genet* 2017;49:1676–83.
- [14] Murthy MN, Blauwendraat C, UK Brain Expression Consortium, Guelfi S, The International Parkinson Disease Genomics Consortium, Hardy J, et al. Increased brain expression of *GPNMB* is associated with genome wide significant risk for Parkinson's disease on chromosome 7p15.3. *Neurogenetics* 2017;18:121–33.
- [15] Kretschmar H. Brain banking: opportunities, challenges and meaning for the future. *Nat Rev Neurosci* 2009;10:70–8.
- [16] Bell JE, Alafuzoff I, Al-Sarraj S, Arzberger T, Bogdanovic N, Budka H, et al. Management of a twenty-first century brain bank: experience in the BrainNet Europe consortium. *Acta Neuropathol* 2008;115:497–507.
- [17] Nichols L, Freund M, Ng C, Kau A, Parisi M, Taylor A, et al. The National Institutes of Health NeuroBioBank: a federated national network of human brain and tissue repositories. *Biol Psychiatry* 2014;75:e21–2.
- [18] Palmer-Aronsten B, Sheedy D, McCrossin T, Kril J. An international survey of brain banking operation and characterization practices. *Biopreserv Biobank* 2016;14:464–9.
- [19] Stan AD, Ghose S, Gao XM, Roberts RC, Lewis-Amezcuca K, Hatanpaa KJ, et al. Human postmortem tissue: what quality markers matter? *Brain Res* 2006;1123:1–11.
- [20] White K, Yang P, Li L, Farshori A, Medina AE, Zielke HR. Effect of postmortem interval and years in storage on RNA quality of tissue at a repository of the NIH NeuroBioBank. *Biopreserv Biobank* 2018;16:148–57.
- [21] Schmitt A, Parlapani E, Bauer M, Heinsen H, Falkai P. Is brain banking of psychiatric cases valuable for neurobiological research? *Clinics (Sao Paulo)* 2008;63:255–66.
- [22] Yan XX, Ma C, Bao AM, Wang XM, Gai WP. Brain banking as a cornerstone of neuroscience in China. *Lancet Neurol* 2015;14:136.
- [23] Zhang H, Chen K, Wang N, Zhang D, Yang Q, Zhang Q, et al. Analysis of brain donors' demographic and medical characteristics to facilitate the construction of a human brain bank in China. *J Alzheimers Dis* 2018;66:1245–54.
- [24] Pipes L, Li S, Bozinovski M, Palermo R, Peng X, Blood P, et al. The non-human primate reference transcriptome resource (NHPTR) for comparative functional genomics. *Nucleic Acids Res* 2013;41:D906–14.
- [25] Doan RN, Bae BI, Cubelos B, Chang C, Hossain AA, Al-Saad S, et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* 2016;167:341–54.e12.
- [26] Samarasekera N, Al-Shahi Salman R, Huitinga I, Klioueva N, McLean CA, Kretschmar H, et al. Brain banking for neurological disorders. *Lancet Neurol* 2013;12:1096–105.
- [27] Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature* 2014;508:199–206.

- [28] Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 2012;489:391–9.
- [29] Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* 2014;17:1418–28.
- [30] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580–5.
- [31] Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 2016;19:1442–53.
- [32] BrainSeq: A Human Brain Genomics Consortium. BrainSeq: neurogenomics to drive novel target discovery for neuropsychiatric disorders. *Neuron* 2015;88:1078–83.
- [33] De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at *ANK1*, *BINI*, *RHBDF2* and other loci. *Nat Neurosci* 2014;17:1156–63.
- [34] PsychENCODE Consortium, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, et al. The PsychENCODE project. *Nat Neurosci* 2015;18:1707–12.
- [35] Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 2011;478:519–23.
- [36] Bennett DA, Schneider JA, Arvanitakis Z, Wilson RS. Overview and findings from the religious orders study. *Curr Alzheimer Res* 2012;9:628–45.
- [37] Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and findings from the rush memory and aging project. *Curr Alzheimer Res* 2012;9:646–63.
- [38] Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 2018;362:eaat7615.
- [39] Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* 2018;362:eaat8464.
- [40] Rajarajan P, Borrmann T, Liao W, Schrode N, Flaherty E, Casino C, et al. Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science* 2018;362:eaat4311.
- [41] Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado-Torres L, et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci* 2018;21:1117–25.
- [42] Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci* 2016;19:40–7.
- [43] Zhu Y, Sousa AMM, Gao T, Skarica M, Li M, Santpere G, et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* 2018;362:eaat8077.
- [44] Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 2018;362:eaat8127.
- [45] Meng Q, Wang K, Brunetti T, Xia Y, Jiao C, Dai R, et al. The *DGCR5* long noncoding RNA may regulate expression of several schizophrenia-related genes. *Sci Transl Med* 2018:eaat6912.
- [46] Chen C, Meng Q, Xia Y, Ding C, Wang L, Dai R, et al. The transcription factor *POU3F2* regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. *Sci Transl Med* 2018:eaat8178.
- [47] Locascio JJ, Atri A. An overview of longitudinal data analysis methods for neurological research. *Dement Geriatr Cogn Dis Extra* 2011;1:330–57.
- [48] Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016;17:13.
- [49] Teng M, Love MI, Davis CA, Djebali S, Dobin A, Graveley BR, et al. A benchmark for RNA-seq quantification pipelines. *Genome Biol* 2016;17:74.
- [50] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28:511–5.
- [51] Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* 2013;10:71–3.
- [52] Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 2010;464:773–7.
- [53] Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016;34:525–7.
- [54] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:323.
- [55] Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol* 2014;32:462–4.
- [56] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017;14:417–9.
- [57] GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* 2017;550:204–13.
- [58] Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* 2018;359:693–7.
- [59] Oldham MC, Langfelder P, Horvath S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst Biol* 2012;6:63.
- [60] Tsoucas D, Dong R, Chen H, Zhu Q, Guo G, Yuan GC. Accurate estimation of cell-type composition from gene expression data. *Nat Commun* 2019;10:2975.
- [61] Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* 2012;7:500–7.
- [62] Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;28:882–3.
- [63] Fusi N, Stegle O, Lawrence ND. Joint modelling of confounding factors and prominent genetic regulators provides increased accuracy in genetical genomics studies. *PLoS Comput Biol* 2012;8:e1002330.
- [64] Gaffney DJ, Veyrieras JB, Degner JF, Pique-Regi R, Pai AA, Crawford GE, et al. Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol* 2012;13:R7.
- [65] van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods* 2015;12:1061–3.
- [66] Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016;48:481–7.
- [67] Ward MC, Gilad Y. Human genomics: cracking the regulatory code. *Nature* 2017;550:190–1.
- [68] Pavlidis JM, Zhu Z, Gratten J, McRae AF, Wray NR, Yang J. Predicting gene targets from integrative analyses of summary

- data from GWAS and eQTL studies for 28 human complex traits. *Genome Med* 2016;8:84.
- [69] Barbeira A, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018;9:1825.
- [70] Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 2015;47:1091–8.
- [71] Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 2016;48:245–52.
- [72] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;10:e1004383.
- [73] Giambartolomei C, Liu JZ, Zhang W, Hauberg M, Shi H, Boockch J, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* 2018;34:2538–45.
- [74] Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *bioRxiv* 2019;592238.
- [75] He X, Fuller CK, Song Y, Meng Q, Zhang B, Yang X, et al. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet* 2013;92:667–80.
- [76] Lee D, Williamson VS, Bigdeli TB, Riley BP, Fanous AH, Vladimirov VI, et al. JEPeG: a summary statistics based tool for gene-level joint testing of functional variants. *Bioinformatics* 2015;31:1176–82.
- [77] Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying causal variants at loci with multiple signals of association. *Genetics* 2014;198:497–508.
- [78] Hormozdiari F, van de Bunt M, Segre AV, Li X, Joo JWJ, Bilow M, et al. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet* 2016;99:1245–60.
- [79] Yang F, Wang J, Consortium GT, Pierce BL, Chen LS. Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis. *Genome Res* 2017;27:1859–71.
- [80] Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 2016;32:1493–501.
- [81] Gamazon ER, Segre AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* 2018;50:956–67.
- [82] Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;47:291–5.
- [83] Aten JE, Fuller TF, Lusk AJ, Horvath S. Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC Syst Biol* 2008;2:34.
- [84] Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018;9:224.
- [85] Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res* 2014;24:14–24.
- [86] Wen X, Pique-Regi R, Luca F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet* 2017;13:e1006646.
- [87] Li X, Kim Y, Tsang EK, Davis JR, Damani FN, Chiang C, et al. The impact of rare variation on gene expression across tissues. *Nature* 2017;550:239–43.
- [88] Diao Y, Fang R, Li B, Meng Z, Yu J, Qiu Y, et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* 2017;14:629–35.
- [89] Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol Ther Nucleic Acids* 2015;4:e264.
- [90] Forrest MP, Zhang H, Moy W, McGowan H, Leites C, Dionisio LE, et al. Open chromatin profiling in hiPSC-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. *Cell Stem Cell* 2017;21:305–18.e8.
- [91] Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, et al. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. *PLoS Genet* 2012;8:e1002982.
- [92] Darnell RB. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* 2010;1:266–86.
- [93] Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008;456:464–9.
- [94] Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, et al. A majority of m⁶A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev* 2015;29:2037–53.
- [95] Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* 2012;150:710–24.
- [96] French JD, Ghoussaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, et al. Functional variants at the 11q13 risk locus for breast cancer regulate *cyclin D1* expression through long-range enhancers. *Am J Hum Genet* 2013;92:489–503.
- [97] Praetorius C, Grill C, Stacey SN, Metcalf AM, Gorkin DU, Robinson KC, et al. A polymorphism in *IRF4* affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* 2013;155:1022–33.
- [98] Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. *FTO* obesity variant circuitry and adipocyte browning in humans. *N Engl J Med* 2015;373:895–907.
- [99] Rhie SK, Coetzee SG, Noushmehr H, Yan C, Kim JM, Haiman CA, et al. Comprehensive functional annotation of seventy-one breast cancer risk loci. *PLoS One* 2013;8:e63925.
- [100] Roussos P, Mitchell AC, Voloudakis G, Fullard JF, Pothula VM, Tsang J, et al. A role for noncoding variation in schizophrenia. *Cell Rep* 2014;9:1417–29.
- [101] Bershteyn M, Nowakowski TJ, Pollen AA, Di Lullo E, Nene A, Wynshaw-Boris A, et al. Human iPSC-derived cerebral organoids model cellular features of lissencephaly and reveal prolonged mitosis of outer radial glia. *Cell Stem Cell* 2017;20:435–49.e4.
- [102] Fan X, Dong J, Zhong S, Wei Y, Wu Q, Yan L, et al. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res* 2018;28:730–45.
- [103] Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* 2016;352:1586–90.
- [104] Aggarwal NT, Bienias JL, Bennett DA, Wilson RS, Morris MC, Schneider JA, et al. The relation of cigarette smoking to incident Alzheimer's disease in a biracial urban community population. *Neuroepidemiology* 2006;26:140–6.

- [105] Wilson RS, Bienias JL, Mendes de Leon CF, Evans DA, Bennett DA. Negative affect and mortality in older persons. *Am J Epidemiol* 2003;158:827–35.
- [106] Sabunciyar S. Gene expression profiles associated with brain aging are altered in schizophrenia. *Sci Rep* 2019;9:5896.
- [107] Viguera AC, Tondo L, Baldessarini RJ. Sex differences in response to lithium treatment. *Am J Psychiatry* 2000;157:1509–11.
- [108] Tohen M, Greil W, Calabrese JR, Sachs GS, Yatham LN, Oerlinghausen BM, et al. Olanzapine versus lithium in the maintenance treatment of bipolar disorder: a 12-month, randomized, double-blind, controlled clinical trial. *Am J Psychiatry* 2005;162:1281–90.
- [109] Rajkowska G, Halaris A, Selemon LD. Reductions in neuronal and glial density characterize the dorsolateral prefrontal cortex in bipolar disorder. *Biol Psychiatry* 2001;49:741–52.
- [110] Bertolino A, Frye M, Callicott JH, Mattay VS, Rakow R, Shelton-Repella J, et al. Neuronal pathology in the hippocampal area of patients with bipolar disorder: a study with proton magnetic resonance spectroscopic imaging. *Biol Psychiatry* 2003;53:906–13.
- [111] Bouras C, Kovari E, Hof PR, Riederer BM, Giannakopoulos P. Anterior cingulate cortex pathology in schizophrenia and bipolar disorder. *Acta Neuropathol* 2001;102:373–9.
- [112] Rajkowska G. Postmortem studies in mood disorders indicate altered numbers of neurons and glial cells. *Biol Psychiatry* 2000;48:766–77.
- [113] Jiao C, Yan P, Xia C, Shen Z, Tan Z, Tan Y, et al. BrainEXP: a database featuring with spatiotemporal expression variations and co-expression organizations in human brains. *Bioinformatics* 2019;35:172–4.