Emerging Methods and Resources for Biological Interrogation of Neuropsychiatric Polygenic Signal

Emil Uffelmann and Danielle Posthuma

ABSTRACT
Most neuropsychiatric disorders are highly polygenic, implicating hundreds to thousands of causal genetic variants that span much of the genome. This widespread polygenicity complicates biological understanding because no single variant can explain disease etiology. A strategy to advance biological insight is to seek convergent functions among the large set of variants and map them to a smaller set of disease-relevant genes and pathways. Accordingly, functional genomic resources that provide data on intermediate molecular phenotypes, such as gene-expression and methylation status, can be leveraged to functionally annotate variants and map them to genes. Such molecular quantitative trait locus mappings can be integrated with genome-wide association studies to make sense of the polygenic signal that underlies complex disease. Other resources that provide data on the 3-dimensional structure of chromatin and functional importance of specific genomic regions can be integrated similarly. In addition, mapped genes can then be tested for convergence in biological function, tissue, cell type, or developmental stage. In this review, we provide an overview of functional genomic resources and methods that can be used to interpret results from genome-wide association studies, and we discuss current challenges for biological understanding and future requirements to overcome them.

https://doi.org/10.1016/j.biopsych.2020.05.022

AVAILABLE RESOURCES TO ASSESS THE POTENTIAL IMPACT OF GENETIC VARIANTS ON GENE FUNCTION
The first step in interpreting GWAS results is to annotate trait-associated variants with their putative impact on gene function (Figure 1). Generally, there are two routes by which a variant may influence disease risk. First, a variant may affect the structure of a protein and lead to disruption of normal functioning and subsequent disease. Here, nonsynonymous exonic variants are most easily interpretable, as the structure of a protein directly relates to its function (3). Variants in coding as well as noncoding regions that affect alternative splicing may have similar consequences by producing protein isoforms (4,5).

Second, a genetic variant may affect the expression level of a gene. Transcriptional control is believed to play a fundamental role in disease (6), and GWAS hits are found to be enriched in regulatory sequences (7). Variants that influence gene expression, known as expression quantitative trait loci (eQTL) experiments. We also address specific challenges related to neuropsychiatric disorders.
(eQTLs), can be identified by integrating genotype with gene-expression data (8). The same approach can be applied to identify variants that affect other molecular phenotypes such as methylation, acetylation, chromatin-conformation features, or splicing. Splicing QTLs can affect expression levels of gene isoforms and, in some cases, show stronger enrichment for disease-associated variants than eQTLs (5,9). Collectively, these regulatory variants are referred to as molecular QTLs (molQTLs) (10). Below, we review the various resources that contain biological information relevant for annotating genetic variants (see also Table 1).

**Positional Mapping**

In the early 2000s, the Human Genome Project completed the sequencing of nearly all 3 billion base pairs that make up human DNA (11). As a follow-up, the Encyclopedia of DNA Elements (ENCODE) project aimed to annotate all functional sequences in the human genome using experimental and computational approaches (12) (Table 1). These large, international, and multidisciplinary projects gave rise to many resources and tools that allow annotating genetic variants and mapping them to genes based on their physical location. Many tools (13–20) provide information about the predicted functional impact of variants (e.g., missense, nonsense, frameshift) and positional information (e.g., exonic, intronic, transcription factor binding site). Positional mapping is particularly useful for variants in coding regions, as it directly links them to genes, but the effects on intermediate molecular phenotypes, such as methylation status and expression levels, in noncoding regions may be missed.

**Figure 1.** Moving from trait-associated variants to convergence on biological function, tissue type, or cell type. Genome-wide association studies (GWASs) identify large numbers of associated variants for polygenic traits. These traits can be mapped to a prioritized set of genes through functional and positional annotation, as well as statistical fine-mapping. Gene-set analysis can test for convergence of prioritized genes in biological function, tissue, or cell type. Developmental stage (not depicted) is ideally taken into account. The aim is to generate genetically informed hypotheses about the likely causal mechanisms underlying the trait of interest. molQTL, molecular quantitative trait locus.
| Resource | Data | Tissue/Cell Type | Donors | Disorder | Age | Access | URL |
|----------|------|-----------------|--------|----------|-----|--------|-----|
| 4D Nucleome (51) | Hi-C, 3C, 4C, 5C, and other technologies to map chromatin interactions | >30 tissues/cell lines | Depends on dataset | Healthy | Adult | Many datasets are freely available. | https://data.4dnucleome.org |
| Allen Brain Atlas Cell Types and Transcriptomics (105–107) | scRNAseq, snRNAseq, RNAseq, microarray | Cortex | 2–8 brains, depending on data type and brain region | Healthy | Adult | Processed data are freely available, raw data can be accessed via dbGaP application. | https://portal.brain-map.org/atlases-and-data/maseq |
| BIOS QTL Browser (112) | cis- and trans-mQTL, cis-eQTLs | Whole blood | 2116 eQTL, 3841 mQTL | Healthy | Adult | QTL mapping results are freely available. | https://genenetwork.nl//biosqtlbrowser/ |
| Blood eQTL Browser (113) | cis- and trans-eQTL | Whole blood | 5311 | Healthy | Adult | QTL mapping results are freely available. | https://genenetwork.nl//bloodeqtlbrowser/ |
| BrainSeq (114) | eQTL, differentially expressed genes | Hippocampus, dorsolateral prefrontal cortex | 495 in phase I, 551 in phase II | Bipolar disorder, major depressive disorder, schizophrenia | Pre- and postnatal donors | Most raw and processed data are freely available; genotype data can be accessed via dbGaP application. | http://eqtl.brainseq.org |
| BrainSpan (24,25) | cis-eQTL | 17 cortical and subcortical structures | 57 | Healthy | Pre- and postnatal donors | Gene-expression data are freely available; genotype data can be accessed via dbGaP application. | http://www.brainspan.org//static/home |
| BrainVar (27) | eQTL | Dorsolateral prefrontal cortex | 176 | Healthy | Pre- and postnatal donors | Data are not released yet. | http://www.brainvar.org |
| CommonMind (46) | cis- and trans-eQTL, caQTL | Dorsolateral prefrontal cortex | 980 | Schizophrenia, bipolar disorder | Adult | Data are in part freely available, in part controlled. | https://www.synapse.org//Synapse:syn2759792/wiki/69613 |
| Depression Genes and Networks Study (DGN) (115) | cis- and trans-eQTL, sQTL | Whole blood | 922 | Healthy | Adult | Significant QTL mapping results are freely available, raw data is available via application. | http://dags.stanford.edu//dgn/ |
| Database of Immune Cell Expression, eQTLs, and Epigenomics (DICE) (39) | cis-eQTL | 13 immune cell types | 91 | Healthy | Adult | Gene-expression and cis-eQTL mapping results are freely available, raw data can be accessed via dbGaP application. | https://dice-database.org//downloads |
| Encyclopedia of DNA Elements (ENCODE) (12) | Variety of assays to define genomic functional elements | >75 different cell and tissue types | Depends on dataset | Healthy | 5 life stages, from embryonic to adult | Data is freely accessible. | https://www.encodeproject.org |
| eQTL Catalogue (116) | eQTL | >20 tissues/cell types | Depends on dataset | n/a | n/a | QTL mapping results are freely available. | https://www.ebi.ac.uk/eqt/l/ |
| eQTLGen (21) | cis- and trans-eQTL | Whole blood | 31,684 | Healthy | Adult | QTL mapping results are freely available. | https://www.eqt/ligen.org |
| Resource | Data | Tissue/Cell Type | Donors | Disorder | Age | Access | URL |
|----------|------|----------------|--------|----------|-----|--------|-----|
| Geuvadis (117) | cis-eQTL, cis-mirQTL | Lymphoblastoid cell line | 462 | Healthy | Adult | All data and results are freely available. | https://www.internationalgenome.org/data-portal/data-collection/geuvadis |
| gnomAD (118) | Whole-exome and whole-genome sequencing | n/a | 71,702 (15,708) genomes in v3.0 (v2.1) | Various adult-onset common diseases | Adult | Processed data are freely available. | https://gnomad.broadinstitute.org |
| GSE87112 (53) | Hi-C contact maps | 14 human tissues | 4 | Healthy | Adult | Hi-C data are freely available. | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112 |
| The Genotype–Tissue Expression Project (GTEx v.8) (23) | cis- and trans-eQTL, sQTL | 54 tissues (13 in the brain) | 948 | Healthy | Adult | Processed data and eQTL mappings are freely available; raw data can be accessed via dbGaP application. | https://gtexportal.org/home/ |
| Multiple Tissue Human Expression Resource (MuTHER) (119) | cis-eQTL | Adipose, lymphoblastoid cell line, skin | 856 twins | Healthy | Adult | cis-eQTL mapping results are freely available. | http://www.muther.ac.uk |
| O’Brien et al. (120) | cis-eQTL | Brain | 120 | Healthy | 12–19 postconception | cis-eQTL mapping results are freely available. | https://doi.org/10.6084/m9.figshare.6881825 |
| PsychENCODE (25,44,45) | molQTL, Hi-C contact maps, RNAseq, and more | Cerebral cortex, hippocampus, amygdala, caudate nucleus, nucleus accumbens, cerebellar cortex | ~2500 | Schizophrenia, bipolar disorder, autism spectrum disorder | Pre- and postnatal donors | molQTL mapping results and Hi-C maps are freely available; RNAseq data and other raw data can be accessed via application. | http://www.psychencode.org |
| Roadmap Epigenomics (121) | Predicted enhancer and promoter regions | All major primary tissues and cell types | Hundreds | Healthy | Adult | All datasets are freely available. | http://www.roadmapepigenomics.org |
| Walker et al. (9) | cis-eQTL, cis-sQTL | Cortex | 233 | Healthy | 14–21 postconception | QTL mappings are freely available; raw data can be accessed via dbGaP application. | https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001900.v1.p1 |
| xQTL (42) | eQTL, mQTL, haQTL (acetylation) | Dorsolateral prefrontal cortex | 494 | Healthy | Adult | QTL mapping results are freely available. | http://mostafavilab.stat.ubc.ca/xqtl/ |

All resources are based on human data. There is considerable overlap between some of these resources: eQTLgen contains data from GTEx, DGN, and BIOS. PsychENCODE contains data from CommonMind, GTEx, and Roadmap Epigenomics. eQTL Catalogue contains data from Geuvadis, BrainSeq, xQTL, and DICE.

3C, chromosome conformation capture; 4C, chromosome conformation capture-on-chip; 5C, chromosome conformation capture carbon copy; caQTL, chromatin accessibility QTL; eQTL, expression QTL; haQTL, acetylation QTL; Hi-C, high-throughput chromosome conformation capture; mirQTL, microRNA QTL; molQTL, molecular QTL; mQTL, methylation QTL; QTL, quantitative trait locus; n/a, not applicable; RNAseq, RNA sequencing; scRNAseq, single-cell RNA sequencing; snRNAseq, single-nucleus RNA sequencing; sQTL, splicing QTL.
Transcriptional Landscapes

Several consortia have generated large microarray and RNA sequencing (RNAseq) datasets to profile gene expression in human tissues (Table 1). These datasets are rich and interesting in their own right, but integration with genotype data allows for subsequent eQTL mapping, in which association studies are performed on the expression levels of each gene (6). Most of these studies focus on the effects of cis-eQTLs that are associated with expression levels of nearby genes (<1 Mb), because they tend to have larger effects on gene expression than trans-eQTLs that are associated with the expression levels of distal genes (>5 Mb) (21). Additionally, focusing on a smaller subset of cis-eQTLs reduces the multiple-testing burden drastically.

The Genotype-Tissue Expression (GTEx) project has generated a wealth of data by applying RNAseq to postmortem tissue of 948 donors across 54 tissues, 13 of which are in the brain (22,23). In this resource alone, it was found that almost all protein-coding genes and approximately two thirds of all long noncoding RNA genes had corresponding cis-eQTLs (22). Many of these were tissue specific. Moreover, GWAS hits were significantly enriched for eQTLs. Thus, the integration of GTEx eQTL mappings with GWAS hits allows for the discovery of tissue-specific effects of trait-associated variants on gene-expression levels. However, because of ancestry-specific linkage disequilibrium (LD), top hits from eQTL analyses cannot simply be overlaid with GWAS hits (see Pinpointing the Most Likely Causal Variants).

While GTEx provides information on spatial transcriptomic regulation, resources such as BrainSpan provide corresponding temporal information (24,25). Gene expression is a stochastic phenomenon that varies across development (24), which is mostly due to shifts in maturity and cell-type fractions in tissues (26). Therefore, both spatial and temporal trajectories need to be considered for relevant eQTL identification. BrainSpan provides RNAseq data on a sample of 57 donors across the lifespan, for 17 (sub)cortical brain structures. BrainVar is a new resource that aims to generate a similar dataset, but with a larger sample comprising 176 donors, yielding more precise spatial and temporal profiling (27). These resources can also be used to impute gene-expression levels in independent datasets and conduct transcriptome-wide association studies (TWASs) on the imputed expression levels. Several tools are available for this purpose (28–33).

Although most resources provide information about healthy tissue, the ultimate purpose of interpreting GWAS results is to delineate disease-underlying mechanisms. Other resources, such as PsychENCODE (34), provide RNAseq and genotype data on large samples of postmortem tissue from control subjects as well as donors with schizophrenia, major depressive disorder, or bipolar disorder. A reverse-genetics approach goes beyond mere base-pair distance (52,53). Furthermore, the emerging picture is that the spatial organization of chromatin is highly variable (54) and that this variability directly relates to changes in gene expression (55).

Similar to other intermediate molecular phenotypes, features of chromatin conformation can also be the basis for QTL mapping. The first such studies have recently been carried out and led to the identification of chromatin QTLs that influence chromatin loop strength (56) and chromatin accessibility (57).

Pinpointing the Most Likely Causal Variants

Functional annotation of trait-associated variants using a variety of resources can be easily conducted with the FUMA webtool (58).
and aids in separating variants with currently no known effect on the expression level or structure of a gene’s product from those with known effect. However, whether the annotated effect is also relevant to the trait still needs to be formally tested. Because of the presence of LD, which is the systematic association of alleles at different loci on the same chromosome (59), most trait-associated variants are likely to be correlated with a causal variant rather than being causally involved themselves.

Apart from using functional annotation, trait-associated variants can also be prioritized based on statistical fine-mapping. The simplest strategies are to select all significantly associated, LD-independent variants in a risk locus based on the lowest association p value and LD pattern in the region, or to conduct a stepwise conditional analysis, searching for the minimal set of variants that explain the association signal in a risk locus. However, the variant with the strongest association is not necessarily the causal variant, because of sample fluctuations in allele frequencies and LD, and because it may be in LD with two or more causal variants. More sophisticated statistical fine-mapping strategies rely on a Bayesian framework that incorporates the pattern of association, LD, and possibly functional information to calculate the posterior probability of being causal for each variant (60). Several tools for statistical fine-mapping are available (61–70), of which FINEMAP (67,68) is the current standard. The recently developed method PolyFun (70) combines functional information on coding, conserved, regulatory, and LD-related annotations with statistical fine-mapping methods employed in FINEMAP (67,68) or SuSiE (69).

An additional strategy to interpret GWAS risk loci is to formally test whether the most likely causal variant is shared between the trait of interest and a second, typically molecular, trait. Such colocalization involves bivariate fine-mapping of a shared locus and can be used alongside TWAS approaches (28–33) that merely test for association with imputed transcription levels. Most methods for colocalization rely on a Bayesian model to calculate the posterior probability that the same variant is causal in both traits. The current state-of-the-art tools are COLOC (71) and eCAVIAR (65), which take summary-level statistics as input and are robust against different modes of inheritance (COLOC) or allelic heterogeneity (eCAVIAR). COLOC, however, assumes 1 causal variant per locus (interpreted as “at least 1”), whereas eCAVIAR allows up to 6 causal variants per locus. Enloc (72) additionally allows assessing enrichment of molQTLs, which improves the accuracy of the evaluation of colocalization. FOCUS (73) combines expression imputation (TWAS) with fine-mapping and colocalization and provides credible sets of genes conditional on the TWAS signal.

The accuracy of colocalization methods depends on the sample sizes of molQTL GWASs, which are currently relatively small. In addition, genetic variants that are not prioritized after colocalization can still be causally involved via other mechanisms and should not be dismissed. However, with increasing molQTL GWAS sample sizes, colocalization methods can aid in prioritizing variants and genes and can provide additional mechanistic insight.

**GENE TO FUNCTION: LOOKING FOR CONVERGENCE**

To provide meaningful starting points for functional follow-up experiments, the polygenic nature of neuropsychiatric disorders must be considered when interpreting the results of GWASs. One way is to look for convergence among the vast number of associated or most likely causal variants. Convergence can be tested using pathway-based association tools that rely on a collection of predefined gene sets and evaluate trait associations of the genes that are part of them (74). Table 2 provides a brief overview of resources that can be used to define gene sets, and this is explained more thoroughly in the Definitions of Biological Functions and Pathways in the Supplement. The interpretability and reliability of such an analysis critically depend on accurate gene annotations of biological functions and pathways, as well as sound statistical methods to evaluate the strength of statistical association above chance.

**Tools for Systematic Evaluation of Convergence of Trait-Associated Variants in Biologically Relevant Functions**

Lists of gene sets can be used to investigate whether trait-associated variants converge on specific gene sets, by conducting pathway or gene-set analysis (GSA). GSA entails evaluating whether genes that are part of the gene set show a generally stronger association with the trait of interest compared with the association of genes that are not part of the gene set. This analysis can be accomplished by using a binary measure of gene-trait association and conducting a $2 \times 2$ hypergeometric test or a quantitative measure of gene-trait association (e.g., transformed $p$ values for association) and using a regression framework.

There are several tools available for GSA (75–85), some of which are reviewed in (74). The most widely used tools are stratified LDSC (86) and MAGMA (84). Although LDSC is not strictly a GSA tool, it does allow for the evaluation of enrichment of single nucleotide polymorphism–based heritability in functional categories of variants. MAGMA (84) relies on allocating variants to genes and uses a multiple regression model that evaluates whether the gene-based association signal inside a gene set is stronger than that outside the gene set. As with any other regression, inclusion of covariates is straightforward, and by default, MAGMA conditions on gene size, gene density, the inverse of the mean minor allele count in each gene, and the number of genes in a gene set. While the default is to assign variants to genes based on a window of 1 kb on each side of the transcription start and end sites, this window can be adjusted. However, there are no agreed-upon standards for window sizes around genes, as each gene may require a different window (87), reflecting an optimal balance between including relevant variants and excluding noise.

Systematic comparisons between different GSA tools and their underlying algorithms have indicated several important issues (Table S1) (84,88–91).

First, statistical power is only marginally related to the sample size of the GWAS and is mostly influenced by the genetic effect per gene (or variant) and the number of genes in a gene set, in addition to the ratio of the heritability due to the genes inside and outside the set. If the set-specific heritability is low (<5%), gene-set sizes of 100 genes still have adequate (80%) power, while gene-set sizes of 25 genes will have low (<10%) statistical power. However, if the gene-set heritability

---

*Biological Psychiatry January 1, 2021; 89:41–53 [www.sobp.org/journal]*
| Resource | URL | Brief Description | Type |
|----------|-----|------------------|------|
| **Resources with Readily Available Lists of Gene Sets Based on Biological Function** | | | |
| DrugBank (122) | [https://www.drugbank.ca](https://www.drugbank.ca) | Provides lists of genes that are targets of specified drugs. Currently contains 13,491 drug entries linked to 5183 unique protein sequences. | Drug targets |
| Gene Ontology (GO) (93) | [http://geneontology.org](http://geneontology.org) | Provides hierarchical lists of genes clustered according to molecular function, cellular location, or biological process. Based on experimental and predicted evidence from 4467 organisms, partly manually curated. Currently contains 44,945 unique annotation terms. | Molecular function, cellular location, biological process |
| Kyoto Encyclopedia of Genes and Genomes - Pathways (KEGG Pathways) (123) | [https://www.kegg.jp/kegg/pathway.html](https://www.kegg.jp/kegg/pathway.html) | Provides lists of genes in pathways based on molecular interaction, reaction, and relation networks, in the context of metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases, and drug development. All maps are manually drawn and based on experimental evidence from a variety of organisms (including 536 eukaryotes, 5518 bacteria, and 313 archaea). | Molecular function |
| Molecular Signatures Database (MSigDB) (124) | [https://www.gsea-msigdb.org/gsea/msigdb/index.jsp](https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) | This is a metadatabase providing information on a collection of gene sets obtained from other resources. | Molecular function, cellular location, biological process, drug targets |
| Reactome (125) | [https://reactome.org](https://reactome.org) | Provides lists of manually curated relations between signaling and metabolic molecules organized into biological pathways and processes. | Biological process |
| SynGO (126) | [https://syngoportal.org](https://syngoportal.org) | Provides lists of genes for 2922 expert-curated functional annotations of 1112 genes. All annotated functions are in the context of the synapse and evidence is derived from cellular experiments based on mouse models. | Synaptic function, cellular location in the synapse, biological process in the synapse |
| WikiPathways (127) | [https://www.wikipathways.org/index.php/WikiPathways](https://www.wikipathways.org/index.php/WikiPathways) | Provides lists of genes, proteins, and metabolites in 2614 biological pathways based on evidence across multiple species. Pathways are curated by approved members from the biological community aided by computer-based algorithms. Currently covers 11,532 human genes. | Molecular function |
| **Resources That Can Be Used to Create Quantitative Gene Sets for Tissue Types** | | | |
| BrainSeq (114) | [http://eqtl.brainseq.org](http://eqtl.brainseq.org) | Provides transcriptome data across distinct brain regions (dorsolateral prefrontal cortex and hippocampus) across the human lifespan and 3 main psychiatric diagnostic groups (schizophrenia, major depressive disorder, bipolar disorder). Allows for creation of sets of differentially expressed genes as well as brain region–specific gene sets. | RNAseq |
| BrainSpan (24,25) | [http://www.brainspan.org/static/home](http://www.brainspan.org/static/home) | Provides transcriptome data across multiple cortical and subcortical human brain regions, at different stages of brain development. | RNAseq |
| Geuvadis (117) | [https://www.internationalgenome.org/data-portal/data-collection/geuvadis](https://www.internationalgenome.org/data-portal/data-collection/geuvadis) | Provides transcriptome data on lymphoblastoid cell lines for 465 individuals from 5 of the populations in the 1000Genomes project. | RNAseq |
| The Genotype–Tissue Expression Project (GTEx v.8) (22,23) | [https://gtexportal.org/home/](https://gtexportal.org/home/) | Provides transcriptome data for 54 human tissues. | RNAseq |
| **Resources That Can Be Used to Create Quantitative Gene Sets for Cell Types** | | | |
| Allen Brain Atlas (105–107) | [https://portal.brain-map.org/atlasthes-and-data/maseq](https://portal.brain-map.org/atlasthes-and-data/maseq) | Provides cell type–specific transcriptome data across several brain regions based on mouse and human samples. | scRNAseq, snRNAseq |
| Broad Single Cell Portal (128) | [https://portals.broadinstitute.org/single_cell/study-drcn-seq-single-nucleus-ma-seq-on-human-archived-brain](https://portals.broadinstitute.org/single_cell/study-drcn-seq-single-nucleus-ma-seq-on-human-archived-brain) | Provides cell type–specific transcriptome data on postmortem human and mouse hippocampal and cerebral cortex samples. | snRNAseq |
Methods and Resources for Interpretation of GWAS Hits

Table 2. Continued

| Resource | URL | Brief Description | Type |
|----------|-----|-------------------|------|
| Database of Immune Cell Expression, eQTLs, and Epigenomics (DICE) (139) | https://dice-database.org/downloads | Provides cell type–specific transcriptome data for mononuclear immune cell types. | RNAseq |
| DropViz (129) | http://dropviz.org/ | Provides cell type–specific transcriptome data for 9 brain regions of mouse samples. | scRNAseq |
| Gene Expression Omnibus (GEO) (130,131) | https://www.ncbi.nlm.nih.gov/geo | Provides cell type–specific transcriptome data for several brain regions and blood, mostly based on mouse samples, with some based on human samples, and including temporal information. | scRNAseq, snRNAseq |
| Linnarsson Lab Mouse Brain Atlas (132) | http://mousebrain.org | Provides cell type–specific transcriptome data in the mouse brain. | scRNAseq |
| The Mouse Cell Atlas (133) | http://bis.zju.edu.cn/MCA/ | Provides cell type–specific transcriptome data across all of the mouse main organs. | scRNAseq |
| Tabula Muris (134) | https://tabula-muris.ds.czbiohub.org | Provides cell type–specific transcriptome data levels across different organs of mice. | scRNAseq |

Resources That Can Be Used to Create Sets of Genes Linked Via Protein Interaction Networks

| Resource | URL | Brief Description | Type |
|----------|-----|-------------------|------|
| Human Integrated Protein–Protein Interaction Reference (HIPPIE) (135) | http://cbdm-01.zdv.uni-mainz.de/~mschaef/hippie | Provides genome-wide information on context-specific PPIs based on integrated experimental data from expression information, GO function and Medical Subject Headings (MeSH) disease terms. Protein–protein annotations are provided with a confidence score indicating how likely it is that the interaction is true. | PPI |
| The Human Reference Interactome and Literature Benchmark (HuRI) (136) | http://www.interactome-atlas.org | Provides information on binary PPIs. PPIs are based on Y2H and validated in multiple orthogonal assays. Currently 17,500 proteins have been tested, and 64,006 interactions involving 9094 proteins have been identified. Also includes a literature-curated subset of currently 13,441 PPIs and 6047 proteins. | PPI |
| The Molecular Interaction Database (MINT) (137) | https://mint.bio.uniroma2.it | Provides information on genome-wide functional PPIs reported in the peer-reviewed literature. Currently contains 26,344 interactors from 647 organisms. | PPI |
| StringDB (138) | https://string-db.org | Provides genome-wide information on PPIs observed or predicted in 5090 organisms for 24.6 million proteins. Most interactions are based on Y2H experiments. Protein–protein annotations are provided with a confidence score indicating how likely it is that the interaction is true. | PPI |

This table does not provide an exhaustive overview of all resources but lists the largest ones that are kept updated. Pathguide (139) provides a more comprehensive list of resources.

PPI, protein–protein interaction; RNAseq, RNA sequencing; scRNAseq, single-cell RNA sequencing; snRNAseq, single-nucleus RNA sequencing; Y2H, yeast 2-hybrid screening.

*Watanabe et al. (38) provide a detailed overview of subdatasets available from these resources.

is around 20%, set sizes of both 100 and 25 have adequate power (92). Second, competitive testing, which tests the null hypothesis that the trait association of genes in the gene set is not stronger than of genes outside it, is almost always preferred over often-biased self-contained testing, which tests the null hypothesis that genes in the gene set are not associated with the trait of interest. This approach is especially important for traits that are known to be polygenic, with nearly all genes showing at least some association with the trait of interest. Third, statistical power of GSA generally decreases as the trait heritability increases and it becomes less likely that most of the genetic variance resides in the gene set of interest. This power loss occurs because with higher trait heritability, the genetic signal is divided across many genes under a polygenic model, increasing the heritability due to genes that are not included in the set of interest. Fourth, the test for statistical association should accurately control for the number of variants, the number of genes, and the level of LD, and some tools have been shown to be more sensitive to these biases than others (92). If the number of genes in a gene set increases, the likelihood of observing a significant association in that gene set by chance also increases.

Genes generally tend to have multiple functions and can be expressed in multiple tissues and cell types (22,23,93). In addition, classes of functional gene properties tend to be correlated across genes. For example, a set of genes involved...
in vesicle docking in the synapse will also show preferential expression in brain tissue. If this set of genes is enriched for strong statistical associations with a trait of interest, traditional GSA testing will result in significant associations for the “vesicle docking” and the “expressed in brain tissue” gene sets, unable to tell which of these is the most likely causal factor. Conditional GSA testing (94) addresses this issue by evaluating the evidence of association of one set (e.g., “vesicle docking”) versus or in combination with another set (e.g., “expressed in brain tissue”). Such post hoc GSA testing has been shown to refine hypotheses on where convergence of polygenic signal resides (95), which is of crucial importance for designing meaningful functional follow-up experiments.

CHALLENGES SPECIFIC TO NEUROPSYCHIATRIC DISORDERS

Although GWASs for neuropsychiatric traits have yielded many replicable hits, several challenges specific to neuropsychiatric disorders may complicate post-GWAS analyses. First, the brain is the most relevant tissue for neuropsychiatric disorders, yet relatively few molQTL resources include extensive information on spatial and temporal characteristics of the brain, and even fewer are disease specific. Many neuropsychiatric disorders manifest in adulthood yet are believed to be developmental in nature, which further impedes access to relevant tissue. Second, even if the whole brain has been mapped in terms of regional and temporal expression of genes, this information will not immediately lead to mechanistic disease insight into neuropsychiatric disorders. GWASs and post-GWAS analyses will merely allow us to formulate hypotheses on convergence of trait-associated variants in specific brain regions, at a specific developmental stage, and potentially for specific cell types. However, functional follow-up experiments need a clear biomarker that can be evaluated in vivo in case-control groups or human brain tissue samples. Therefore, more extensive GWAS-informed fundamental biological experiments are needed that map the temporal and spatial circuitry of specific cell types and how they are affected by genetic variation, while simultaneously dealing with the polygenic nature of neuropsychiatric disorders. Tools such as chemogenetic manipulation with DREADDs (designer receptors exclusively activated by designer drugs) (96) or induced pluripotent stem cell technology (97,98) that do not rely on manipulating single genes are particularly relevant in this regard. Third, neuropsychiatric disorders show extensive genetic correlations (99), which may be due to diagnostic (symptom) overlap, pleiotropic effects of causal genes, or causal relations between multiple traits. Thus, the study of a specific neuropsychiatric trait requires considering its pheno- and genotypic overlap with other neuropsychiatric traits. Genetic overlap can be quantified with LDHub (100), which automates and scales LDSC (85) genetic correlation analyses, or GCTA (101) when raw genotypic data are available. GenomicSEM (102) allows for more sophisticated modelling of relationships between multiple traits.

CONCLUSIONS

The widespread polygenicity of neuropsychiatric disorders precludes the identification of single disease-relevant genes and complicates the formulation of mechanistic hypotheses that can be tested in functional follow-up studies. To interpret outcomes from GWASs in biological context requires 1) functional annotation of trait-associated variants using external biological resources, 2) prioritization of likely causal variants and genes, and 3) evaluation of convergence at multiple biological levels using appropriate statistical methods. We have provided an overview of currently available functional genomic resources and methods that can be used in each of these 3 steps. There are several ways in which the currently available resources and methods can be optimized.

First, most resources for annotation of the regulatory function of trait-associated variants relies on transcriptional data, while other intermediate molecular phenotypes may also be required to explain the regulatory consequences of genetic variation. More and larger studies are needed that map multiple molQTLs, preferably in an integrated manner, such as is done by the PsychENCODE (34) consortium. Second, current resources for single-variant annotation are mostly based on adult postmortem tissue, disregarding the temporal variation in molecular phenotypes. Temporal variation in gene function is especially important for many neuropsychiatric disorders, such as autism spectrum disorder and attention-deficit/hyperactivity disorder, that are developmental in nature (103). As a consequence, relevant biological mechanisms might only become noticeable at certain ages. Additionally, adult tissue may not contain the disease-causing cell types, which may be found exclusively in early developmental stages (104). Resources containing information at multiple developmental stages are thus needed. Third, complementing coarse bulk tissue data with fine-grained single-cell data constitutes another much-needed advance. In the future, the Single-Cell eQTLgen Consortium (https://www.eqtlgen.org/single-cell.html) aims to identify the specific cell types in which eQTL effects emerge (21). The Allen Brain Institute has undertaken the Herculean task of defining all cell types in the adult human and mouse cerebral cortex according to their morphological, electrophysiological, and transcriptional properties (105–107). Human fetal neocortical cell types are now also being mapped comprehensively (108). Taking a step forward, the Human Cell Atlas (109) and HuBMAP (110) aim to define all human cell types. Once genetic variants have been mapped to genes, this information may allow for the delineation of disease mechanisms that are cell type specific. Fourth, relatively few resources provide data on donors with neuropsychiatric disorders. As the ultimate goal is to delineate disease mechanism, contrasting healthy and diseased tissue is likely to provide important insights. Fifth, some resources constitute a patchwork of small studies that impede easy integration with GWASs. In contrast, resources such as GTEx (23) and xQTL (42) provide single and large datasets that are easily accessible and readily integrated. Preferably larger, more homogeneous resources will be available in the future. Lastly, novel methods are needed that integrate information from multiple molQTLs to optimize prioritization of causal variants and genes, and that allow evaluating the effects of variants across the continuous allelic frequency spectrum (for a discussion of rare variants, see the Supplement).

Only by combining spatial and temporal trajectories, diseased and healthy tissue, and bulk and single-cell data and testing for convergence in well-defined gene sets will we be able to paint a
comprehensive picture of our 3-dimensional genomic and epigenomic landscape in the context of neuropsychiatric disorders. But this is only the first step toward true biological insight of disease. The approaches mentioned here are purely exploratory and will only be able to generate hypotheses, as opposed to confirm them. Computational approaches that predict how genetic mutations affect neuronal function at protein, cellular, or network level can be employed to formulate hypotheses that need to be validated in experimental studies. These computational approaches can include the annotated effects of hundreds of genetic variants taking into account the polygenic nature of mental disorders (11). Such computational approaches have successfully been applied in the context of cardiac traits, although they may prove to be more challenging to apply to neuropsychiatric disorders. Hypothesized biological mechanisms are not guaranteed to be immediately relevant for the trait of interest as, for example, the effects of variants may be mediated by environmental mechanisms or the relevant biological pathway may not have been tested. To prove causal involvement of biological mechanisms in neuropsychiatric disorders, we need functional in vitro and in vivo follow-up experiments. This approach requires close collaboration between human geneticists and neurobiologists, and the application of novel neurobiological tools that appropriately deal with the polygenic nature of neuropsychiatric disorders.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was funded by The Netherlands Organization for Scientific Research (Grant No. NWO VICI 435–14–005 [to DP] and NWO Gravitation: BRAINSCAPES: A Roadmap from Neurogenetics to Neurobiology (Grant No. 024.004.012 [to DP]), a European Research Council advanced grant (Grant No. ERC-2018-AdG GWAS2FUNC 834057 [to DP]), and a VU University research fellowship (to EU).

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Complex Trait Genetics (EU, DP), Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam; and the Department of Child and Adolescent Psychiatry and Pediatric Psychology (DP), Section Complex Trait Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam University Medical Center, Amsterdam, The Netherlands. Address correspondence to Danielle Posthuma, Ph.D., at d.posthuma@vu.nl.

Received Feb 19, 2020; revised Apr 23, 2020; accepted May 14, 2020. Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2020.05.022.

REFERENCES

1. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017): 10 years of GWAS discovery: Biology, function, and translation. Am J Hum Genet 101:5–22.
2. Rees E, Owen MJ (2020): Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. Genome Med 12:43. Available at: https://doi.org/10.1186/s13037-020-00734-5. Accessed June 23, 2020.
3. Berg JM, Tymoczko JL, Gatto GJ Jr, Stryer L (2015): Biochemistry, 8th ed. New York: W.H. Freeman & Company.
4. Garcia-Blanco MA, Baranikaj AP, Lasda EL (2004): Alternative splicing in disease and therapy. Nat Biotechnol 22:535–546.
5. Li Y, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, et al. (2016): RNA splicing is a primary link between genetic variation and disease. Science 352:600–604.
6. Lee TI, Young RA (2013): Transcriptional regulation and its misregulation in disease. Cell 152:1237–1251.
7. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. (2012): Systematic localization of common disease-associated variation in regulatory DNA. Science 337:1190–1195.
8. Gilad Y, Ritiksa PA, Pritchard JK (2008): Revealing the architecture of gene regulation: The promise of eQTL studies. Trends Genet 24:408–415.
9. Walker RL, Ramaswami G, Hartl C, Mancuso N, Gandal MJ, de la Torre-Ubieta L, et al. (2019): Genetic control of expression and splicing in developing human brain informs disease mechanisms. Cell 179:750–771, e22.
10. Delaneau O, Ongen H, Brown AA, Fort A, Panouissis NI, Dermitzakis ET (2017): A complete tool set for molecular QTL discovery and analysis. Nat Commun 8:15452.
11. International Human Genome Sequencing Consortium (2004): Finishing the euchromatic sequence of the human genome. Nature 431:931–945.
12. The ENCODE Project Consortium (2012): An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74.
13. Wang K, Li M, Hakonarson H (2010): ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164–e164.
14. Ng PC, Henikoff S (2003): SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812–3814.
15. Liu X, Wu C, Li C, Boerwinkle E (2016): dbNSFP v3.0: A one-stop database of functional predictions and annotations for human non-synonymous and splice-site SNVs. Hum Mutat 37:235–241.
16. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J (2014): A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46:310–315.
17. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. (2010): A method and server for predicting damaging missense mutations. Nat Methods 7:248–249.
18. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Altfoeldi J, Wang Q, et al. (2020): The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581:434–443.
19. McLaren W, Gil L, Hunt SE, Ritchie GS, Thorann A, et al. (2016): The Ensembl Variant Effect Predictor. Genome Biol 17:122.
20. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S (2010): Identifying a high fraction of the human genome to be under selective constraint using GERP++-r. PLoS Comput Biol 6:e1001025.
21. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. (2018): Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv. https://doi.org/10.1101/447367.
22. Aguet F, Barbera AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. (2019): The GTEx consortium atlas of genetic regulatory effects across human tissues. bioRxiv. https://doi.org/10.1101/78903.
23. Lonsdale J (2013): The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580–585.
24. Kang HJ, Kawasaki YI, Cheng F, Zhu Y, Xu X, Li M, et al. (2011): Spatio-temporal transcriptome of the human brain. Nature 478:483–489.
25. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, et al. (2014): Transcriptional landscape of the prenatal human brain. Nature 508:199–206.
26. Wang D, Liu S, Warell J, Won H, Shi X, Navarro FCP, et al. (2018): Comprehensive functional genomic resource and integrative model for the human brain. Science 362:aat8464.
27. Werling DM, Pochareddy S, Choi J, An J-Y, Sheppard B, Peng M, et al. (2019): BrainVar data set: Whole-genome and RNA sequencing
Methods and Resources for Interpretation of GWAS Hits

reveal variation and transcriptomic coordination in the developing human prefrontal cortex. Cell Rep 31:107489.

28. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. (2016): Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 48:245–252.

29. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-McAteer K, Carroll RJ, et al. (2015): A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 47:1091–1098.

30. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. (2018): Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 9:1825.

31. Nagpal S, Meng X, Epstein MP, Tsio LC, Patrick M, Gibson G, et al. (2019): TIGAR: An improved Bayesian tool for transcriptomic data imputation enhances gene mapping of complex traits. Am J Hum Genet 105:258–266.

32. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. (2016): Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 48:481–487.

33. Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, et al. (2019): A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet 51:568–576.

34. Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, et al. (2015): The PsychENCODE project. Nat Neurosci 18:1707–1712.

35. Hernandez LM, Kim M, Hofman GD, Haney JR, de la Torre-Ubieta L, Pasanisi B, Gandal MJ (2021): Transcriptomic insight into the polygenic mechanisms underlying psychiatric disorders. Biol Psychiatry 89:65–75.

36. Eberwine J, Sul J-Y, Bartfai T, Kim J (2014): The promise of single-cell sequencing. Nat Methods 11:25–32.

37. Scala F, Kobak D, Bernabucci M, Bernaerts Y, Cadwell CR, Castro JR, et al. (2020): Phenotypic variation within and across transcriptomic cell types in mouse motor cortex. bioRxiv.https://doi.org/10.1101/2020.01.13.929158.

38. Gouwens NW, Sorensen SA, Baftizadeh F, Budzillo A, Lee BR, Bottolo L, Chadeau-Hyam M, Bernabucci M, Bernaerts Y, Cadwell CR, Castro JR, et al. (2020): Phenotypic variation within and across transcriptomic cell types in mouse motor cortex. bioRxiv.https://doi.org/10.1101/2020.02.03.932244.

39. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. (2016): Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 48:245–252.

40. Gouwens NW, Sorensen SA, Baftizadeh F, Budzillo A, Lee BR, Bottolo L, Chadeau-Hyam M, Bernabucci M, Bernaerts Y, Cadwell CR, Castro JR, et al. (2020): Phenotypic variation within and across transcriptomic cell types in mouse motor cortex. bioRxiv.https://doi.org/10.1101/2020.02.03.932244.
69. Wang G, Sarkar A, Carbonetto P, Stephens M (2019): A simple new approach to variable selection in regression, with application to genetic fine-mapping. bioRxiv. https://doi.org/10.1101/501114.

70. Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, et al. (2019): Functionally-informed fine-mapping and polygenic localization of complex trait heritability. bioRxiv. http://doi.org/10.1101/807792.

71. Giambartolomei C, Vulcovic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V (2014): Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10:e1004383.

72. Wen X, Pique-Regí R, Luca F (2017): Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. PLoS Genet 13:e1006646.

73. Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, Pasanici B (2019): Probabilistic fine-mapping of transcriptome-wide association studies. Nat Genet 51:675–682.

74. Wang K, Li M, Hakonarson H (2010): Analysing biological pathways in genome-wide association studies. Nat Rev Genet 11:843–854.

75. Li M-X, Kwan JSH, Shum PC (2012): HYST: A hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis. Am J Hum Genet 91:478–488.

76. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. (2007): PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.

77. Moskvina V, O’Dushlaine C, Purcell S, Craddock N, Holmans P, O’Donovan MC (2011): Evaluation of an approximation method for assessment of overall significance of multiple-dependent tests in a genomewide association study. Genet Epidemiol 35:861–866.

78. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, et al. (2019): Probabilistic approach to variable selection in regression, with application to genome-wide genetic association analysis. bioRxiv. https://doi.org/10.1101/669179.

79. Bender D, Pettersson U, Holmans P, Purcell SM, Sklar P, et al. (2009): Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. Am J Hum Genet 85:13–24.

80. Lee PH, O’Dushlaine C, Thomas B, Purcell SM (2012): INRICH: Interval-based enrichment analysis for genome-wide association studies. Bioinformatics 28:1797–1799.

81. Pedrero I, Lourdusamy A, Rietelsiel M, Nöthen MM, Cichon S, McQuin P, et al. (2012): Common genetic variants and gene-expression changes associated with bipolar disorder are over-represented in brain signaling pathway genes. Biol Psychiatry 72:311–317.

82. Holmans P, Green EK, Pahwa JS, Ferreira MAR, Purcell SM, Sklar P, et al. (2009): Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. Am J Hum Genet 85:13–24.

83. Lee PH, O’Dushlaine C, Thomas B, Purcell SM (2012): INRICH: Interval-based enrichment analysis for genome-wide association studies. Bioinformatics 28:1797–1799.

84. Segre AV, et al. (2010): MAGIC Consortium, MAGIC investigators, Groop L, Mootha VK, Daly MJ, Altshuler D (2010): Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet 6:e1001058.

85. Wang K, Li M, Buican M (2007): Pathway-based approaches for analysis of genomewide association studies. Am J Hum Genet 81:1278–1283.

86. Lee PH, O’Dushlaine C, Thomas B, Purcell SM (2012): INRICH: Interval-based enrichment analysis for genome-wide association studies. Bioinformatics 28:1797–1799.

87. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015): MAGMA: Generalized gene-set analysis of GWAS data. PLoS Comput Biol 11:e1004219.

88. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. (2015): LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295.

89. ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, The RACI Consortium, Finucane HK, Bulik-Sullivan B, Gusev A, et al. (2016): Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47:1228–1235.

90. Hong C-C, Ramlawi JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, et al. (2017): An atlas of human long non-coding RNAs with accurate 5’ ends [no. 7644]. Nature 543:199–204.

91. Cantor RM, Lange K, Sinhasen JS (2010): Prioritizing GWAS results: A review of statistical methods and recommendations for their application. Am J Hum Genet 86:6–22.

92. Hing M-G, Pawitan Y, Magnússon PKE, Prince JA (2009): Strategies and issues in the detection of pathway enrichment in genome-wide association studies. Hum Genet 126:289–301.

93. Kraft P, Raychaudhuri S (2009): Complex diseases, complex genes: Keeping pathways on the right track. Epidemiology 20:508–511.

94. Ebbert CC, van Eijk KR, Franke L, Mulder F, van der Schouw YT, Wijmenga C, Orland-Moret NC (2009): Using genome-wide pathway analysis to unravel the etiology of complex diseases. Genet Epidemiol 33:419–431.

95. de Leeuw CA, Neale BM, Heskes T, Posthuma D (2016): The statistical properties of gene-set analysis. Nat Rev Genet 17:353–364.

96. Ashburner M, Ball CA, Blake JD, Botstein D, Butler H, Cherry JM, et al. (2000): Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29.

97. de Leeuw CA, Stringer S, Dekkers IA, Heskes T, Posthuma D (2018): Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. Nat Commun 9:3768.

98. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. (2018): Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet 50:390–400.

99. Armbuster BN, Li X, Pausch MH, Herlitze S, Roth BL (2007): Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. Proc Natl Acad Sci U S A 104:5163–5168.

100. Brennand KJ, Simone A, Tran N, Gage FH (2012): Modeling psychiatric disorders at the cellular and network levels. Mol Psychiatry 17:1239–1253.

101. Takahashi K, Yamankaya S (2006): Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676.

102. Brainstorm Consortium, Aittila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. (2018): Analysis of shared heritability in common disorders of the brain. Science 360:aap8757.

103. Zheng J, Erurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. (2017): LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33:272–279.

104. Yang J, Lee SH, Goddard ME, Visscher PM (2011): GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82.

105. Grotzinger AD, Rhemtulla M, Vlaming R de, Ritchie SJ, Mallard TT, Hill WD, et al. (2019): Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nat Hum Behav 3:513–525.

106. Harris JC (2014): New classification for neurodevelopmental disorders in DSM-5. Curr Opin Psychiatry 27:95–97.

107. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, Geschwind DH (2018): The dynamic landscape of open chromatin during human cortical neurogenesis. Cell 172:289–304, e18.

108. Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, et al. (2016): Adult mouse cortical cell taxonomy revealed by single cell transcriptionomics. Nat Neurosci 19:335–348.

109. Tasic B, Yao Z, Grayduck LT, Smith KA, Nguyen TN, Bertagnoli D, et al. (2018): Shared and distinct transcriptomic cell types across neocortical areas. Nature 563:72–78.

110. Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. (2019): Conserved cell types with divergent features in human versus mouse cortex. Nature 573:61–68.

111. Poloudakis D, de la Torre-Ubieta L, Langemak J, Eldens AG, Shi X, Stein JL, et al. (2019): A single-cell transcriptomic atlas of human neocortical development during mid-gestation. Neuron 103:785–801, e8.

Methods and Resources for Interpretation of GWAS Hits
Methods and Resources for Interpretation of GWAS Hits

109. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. (2017): The Human Cell Atlas. Elev 6x:27041.
110. HuBMAP Consortium (2019): The human body at cellular resolution: The NIH Human Biomolecular Atlas Program. Nature 574:187–192.
111. Mičí-Kačmazová T, Kaufmann T, Elsvågshagen T, Devor A, Djurovic S, Westby LT, et al. (2019): Biophysical psychiatry—how computational neuroscience can help to understand the complex mechanisms of mental disorders. Front Psychiatry 10:534.
112. Bonder MJ, Lijnik R, Zemackova DV, Moed M, Deelen P, Vermaat M, et al. (2017): Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet 49:131–138.
113. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. (2019): The human body at cellular resolution: The NIH Human Biomolecular Atlas Program. Nature 518:317–323.
114. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. (2014): Methods and Resources for Interpretation of GWAS Hits (MsigDB) 3.0. Bioinformatics 27:1739–1740.
115. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. (2014): The Reactome pathway knowledgebase. Nucleic Acids Res 42:D472–D477.
116. Koopmans F, van Nierop P, Andres-Alonso M, Bynes A, Cijouw T, Coba MP, et al. (2019): SnpGEO: An evidence-based, expert-curated knowledge base for the synapse. Neuron 103:217–234, e4.
117. Schubert CR, O’Connell DM, Jenkins L, Chukwudumere C, Cooper GC, et al. (2013): Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 45:1238–1243.
118. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, Kerimov N, Hayhurst JD, Manning JR, Walter P, Kolberg L, McCormick C, Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Li Z, Cai M, Wu Q, et al. (2020): A reference map of the human binary protein interactome. Cell 174:1015–1030, e16.
119. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. (2013): NCBI GEO: Archive for functional genomics data sets—update. Nucleic Acids Res 41:D991–D995.
120. Harb C, Domrachev M, Lash AE (2002): Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210.
121. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Hoen PAC, et al. (2017): Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 45:1238–1243.
122. Edgar R, Domrachev M, Lash AE (2002): Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210.
123. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP (2011): Molecular signatures database (MsigDB) 3.0. Bioinformatics 27:1739–1740.
124. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. (2019): The human body at cellular resolution: The NIH Human Biomolecular Atlas Program. Nature 574:187–192.