Specific Aims

Psychiatric disorders are often highly heritable. Consequently, mapping susceptibility genes is a proven strategy for understanding the etiology of these traits and identifying potential therapeutic targets. Genome-wide association studies (GWAS) are currently the main paradigm for gene mapping in complex diseases, but their success in psychiatric diseases has been uneven. Notably, GWAS have had limited success in early-onset diseases such as autism. Furthermore, even for loci associated with disease, it has proven difficult to translate these findings into understandings of the underlying mechanisms. Another strategy emerged recently for psychiatric disorders: identifying *de novo* mutations from exome or genome sequencing of affected families. This approach uses an excess of *de novo* mutations in cases as a signature of risk genes. Nevertheless, we lack a statistical framework that integrates an enormous amount of, and highly diverse, genetic variation from sequencing data.

The goal of this proposal is to develop novel and powerful statistical methods for unraveling the genetic basis of complex psychiatric diseases. Existing methods for gene mapping typically analyze one type of data per time. We take a highly integrative approach. Our key rationales are: first, a risk gene of a disease may be genetically perturbed in different ways, e.g. by variants changing amino acid sequences, by large genomic deletions, or by mutations disrupting regulatory elements and consequently expression level. Integrating information from these different types of events targeting the same genes would lead to much higher power for gene discovery. Second, functional effects of genetic variants can be assessed independently of phenotypes, e.g. by epigenomic annotations or by the effects on gene expression. Leveraging these resources jointly will lead to discoveries that would be impossible with genetic data (genotypes and phenotypes) alone. Specifically, we will:

Aim 1. Develop new statistical methods for mapping disease genes from sequencing data on parent-child trios. Trio-sequencing has shown great promise in a number of psychiatric diseases, such as autism and epilepsy. Current analysis approaches, however, are limited, typically making use of only the most severe forms of *de novo* mutations. Less severe, but much more numerous *de novo* missense mutations, as well as inherited variants from parents, are often not utilized. Furthermore, no method currently exists for exploiting non-coding variants from whole genome sequencing, which is quickly becoming affordable. We will develop more powerful methods that make full use of the wide range of variation identified by sequencing. Our methods will also integrate additional biological information from genomic annotations – such as histone marks, sequence conservation and allele frequencies – to further improve power and better prioritize functional variants for gene mapping.

Aim 2. Develop methods that leverage copy number variations (CNVs) for gene mapping. CNVs have been implicated in many neurodevelopmental disorders, and evidence has started to emerge that CNVs and single-nucleotide variants (SNVs) converge on the same set of genes. The challenge is that CNVs often span multiple genes, and it is often unclear which one is the susceptibility gene. Current analysis methods cannot effectively combine CNV data with other variants. We will develop a principled statistical framework to integrate, at the gene-level, SNV and CNV data. Our approach will accommodate both trio-sequencing and case-control studies. The result will be higher sensitivity for detecting risk genes than using either SNVs or CNVs alone.

Aim 3. Develop methods for integrated analysis of GWAS and expression QTL. More than 90% of loci found in GWAS are located in non-coding regions. Expression QTL data expose the molecular consequences of genetic variants in terms of expression changes. Thus linking GWAS of diseases and eQTL data from specific tissues or conditions has great potential to reveal risk genes and the contexts where they act. Nevertheless, existing methods for joint eQTL and GWAS analysis are often limited to only *cis*-eQTL, and cannot distinguish causality of gene-phenotype relation from correlation due to shared genetic loci. We will develop a statistical framework, using summary statistics from GWAS and eQTL studies, to test if a gene has a causal effect on a phenotype. The method uses both *cis*- and *trans*-eQTL of a gene across the entire range of effect sizes, and will be designed to distinguish causal genes from alternative scenarios. The use of *trans*-eQTL, in particular, will enable us to identify novel disease genes that would be impossible to find with GWAS alone.

Central to our efforts is the implementation of **open-source, modular software** for use by the psychiatric genetics community. We will continue our successful collaborations with autism experts to study genetics of autism, and also initiate new collaborations with scientists from the Psychiatric Genomics Consortium (PGC).

Research Strategy

A. Significance

GWAS have detected thousands of loci associated with a range of diseases, including some psychiatric disorders [1, 2, 3, 4]. Despite these successes, GWAS have some major limitations. Most GWAS loci have small effect sizes, reflecting natural selection against deleterious mutations [5, 6]. For early-onset diseases such as autism, risk alleles are generally kept at low frequencies, making them difficult to identify with GWAS [7]. Furthermore, causal variants or genes in GWAS loci are often unclear. The overarching goal of this proposal is to overcome these limitations by developing statistical tools that leverage additional types of data, including sequencing data from families, copy number variations and genetic data of molecular traits (particularly expression QTL).

Sequencing studies of parent-child trios: *De novo* mutations arise spontaneously in offspring, and can be identified by sequencing patient families, usually parent-child trios ("trio-sequencing"). Analyzing *de novo* mutations is a powerful new approach to mapping risk genes, based on the notion that a risk gene harbors more *de novo* mutations in patients than expected by chance [8, 9]. This approach complements GWAS in two important aspects: identification of actual causal genes, and much larger effect sizes because of weak natural selection on new mutations. It has been very successful in a range of diseases including autism [10, 11], intellectual disability [12], schizophrenia [13], epilepsy [14] and congenital heart disease [15]. In the case of autism, trio-sequencing studies identified nearly 70 risk genes, most of which were not clearly linked to autism previously [10, 11, 16].

While trio-sequencing offers great promise, existing statistical methods are far from optimal. Most current work focuses on *de novo* loss-of-function mutations (LoF, often defined as nonsense mutations and frameshift indels), which constitute only 5% of all *de novo* mutations in exome [17]. This misses information in less damaging but much more numerous missense mutations, as well as all inherited variants from parents. An even bigger challenge lies in non-coding variants from whole-genome sequencing (WGS) data. Despite the importance of non-coding variants in complex diseases [18], they were effectively discarded in two recent WGS studies of autism [19, 20]. We believe integrating all these forms of genetic variations – of potentially different effect sizes but targeting the same genes – would greatly improve our ability to map disease genes.

Copy number variations (CNVs): CNVs are large genomic insertion or deletion events, often at the scale of kilo- to mega-bases. While individually rare, CNVs are collectively common and explain ten times or more nucleotide difference between individuals than SNVs [21]. CNVs are an important source of genetic variation affecting neuron-psychiatric disorders. It is estimated that 10% of individuals with developmental delay carry at least one CNV > 500 kb [21]. *De novo* CNVs, and to a lesser extent inherited CNVs, are enriched in patients with psychiatric diseases such as autism [22, 23, 16]. A number of individual CNVs have been associated with multiple neuropsychiatric disorders [21]. Computational methods have been developed to call CNVs from sequencing data, creating great opportunities for CNV analysis following DNA sequencing [24].

Existing approaches for analyzing CNVs are focused either on detection of CNVs associated with disease risks; or identification of gene sets with CNV burden (e.g. deletions affecting a gene set are more common in cases than in controls) [21, 25]. In neither case do researchers obtain direct knowledge of susceptibility genes as CNVs often span multiple genes (1-30 or more) [21]. This represents a very serious limitation of CNV analyses, one that we propose to address here. Our goal is to develop a statistical framework to analyze CNV data at the level of genes so that the results can be easily interpreted and combined with SNV-based studies. Recent studies have shown that SNVs and CNVs of psychiatric diseases may converge on the same set of target genes [26, 27], making joint SNV-CNV analysis an extremely attractive strategy.

Integrating GWAS and expression quantitative trait loci (eQTL): The majority of disease-associated variants (estimated to be as high as 90%) are located in noncoding sequences, potentially altering gene expression rather than protein function [18]. A key step towards understanding the functions of variants discovered from GWAS is thus to profile their effects on gene expression. Indeed GWAS hits are highly enriched with loci associated with gene expression levels (eQTL) and vice versa [28, 29]. Intersecting eQTL and GWAS data has helped identify risk genes in a number of studies [30, 31]. Many efforts, including the large Genotype Tissue Expression project (GTEx), have been undertaken to create a systematic map of eQTL across major human tissues, providing many opportunities for integrated GWAS and eQTL analysis [29]. Nevertheless, most current efforts are based on the simple idea of searching for SNPs associated with both phenotype and gene expression, one SNP a time.

The problem we address is: can we test the role of a gene on a disease by harnessing all eQTL of the gene,

both *in cis* (close to the gene) and *in trans* (distal or different chromosome)? We argue that genes are more meaningful units for analysis than individual variants: they are more interpretable than SNPs, whose targets are often unknown; and there are far fewer genes than variants, reducing the burden of multiple testing. While a few methods exist for integrating eQTL and GWAS at the gene-level, they have some major limitations [32, 33, 34]. Almost all of them focus on *cis*-eQTL, but studies have shown that the majority of expression heritability (60-75%) lie in *trans*-eQTL [35, 36]. And most existing work was designed to uncover some form of correlation between expression levels and phenotypes, namely sharing of genetic loci, rather than establish causality.

We propose to develop **a comprehensive set of tools for gene mapping by integrating multiple types of genetic variations and functional genomic data**. These tools will be tailored and are particularly important for psychiatric genetics. We envision that our proposed work will have the following impacts:

Providing a statistical framework for the analysis of exome and genome sequencing data from families (Aim 1). Our methods will take advantage of all forms of genetic variations, including both coding and noncoding variants, either arising from *de novo* mutations or inherited from parents. We will also exploit annotations of variants from external resources, such as allele frequencies and epigenomic information, to prioritize more deleterious variants. The result will be much more powerful statistical methods for mapping risk genes than existing ones. In addition, the methods will discover specific regulatory elements underlying psychiatric diseases. Because regulatory elements are often active in specific spatial-temporal contexts (e.g. a particular brain region), such findings can provide valuable insights to the disease mechanisms [37, 38].

Providing a principled statistical method for joint analysis of SNV and CNV data (Aim 2). Our basic rationale is: even if we cannot conclusively identify risk genes from CNVs, we can make useful probabilistic statements. For example, suppose we have identified a disease-associated CNV that covers two genes. We now know that each of the two genes have 50% probability of being risk genes, and adding even a small amount of additional data, we may be able to resolve the true risk gene. The approach we develop, using the language of Bayesian statistics, will allow researchers to obtain such probabilistic statements of genes from CNV data (both *de novo* and case-control) and easily combine them with SNV studies.

Providing a novel statistical approach to gene discovery by integrated analysis of eQTL and GWAS (Aim 3). The method we develop identifies disease genes by testing if the entire set of eQTL of a gene, are collectively associated with the disease risk in GWAS. By using *trans*-eQTL, our approach has the potential of discovering new disease genes that display no association signal in its neighborhood, thus would be missed by GWAS [39, 40]. Another key feature of our method is that it aggregates information of both strong (the ones passing significance threshold) and weak SNPs. This leads to increased power as complex traits are often polygenic, with variants of strongest effects contributing only a small percent of heritability [41, 42]. Our method will be designed to broadly applicable: it operates entirely on summary statistics, and can handle other kinds of molecular QTL data. To maximize its utility, we will curate publicly available eQTL and build a web-based system where a user can easily run our software on his/her GWAS data against our eQTL collection.

Advancing our knowledge of autism and schizophrenia. Autism Spectrum Disorder (ASD) affects about one in 68 children in US [43]. Building on our productive collaborations with autism geneticists, we will evaluate and apply the methods to sequencing data from Autism Sequencing Consortium and Simons Simplex collection (letters from Buxbaum and Sanders). Schizophrenia is another fertile area for applying the methods we propose. GWAS, trio-sequencing and CNV studies have all been successful in schizophrenia [4, 13, 25, 27]. It has also been shown that brain eQTL are enriched with schizophrenia loci [44, 45]. We will start new collaboration with investigators of the PGC to explore the genetics of schizophrenia (Sullivan letter).

Creating open-source software. The software we implement will address analytic challenges often encountered in GWAS and sequencing studies, and will have features that make it easy to use such as detailed documentation and modular design. We will work closely with geneticists to obtain feedback on our software.

B. Innovation

Most existing statistical genetic methods were designed to analyze one type of data a time, such as GWAS or CNVs. Our proposal takes a highly integrative approach, combining multiple types of genetic and genomic data. Specifically, **Aim 1 provides a strong statistical foundation for trio-sequencing studies**. The method in Aim 1a allows us to map susceptibility genes by combining information from all mutations related to a gene (both

coding and non-coding, arising *de novo* or inherited), in an optimal way. The variants are automatically weighted by the method, using external annotations that are informative of their likely effects. Aim 1b further improves the power of this approach by taking advantage of clustering of functional variants into natural units, e.g. exons for coding and enhancers of non-coding sequences. Aim 1c proposes a pathway analysis method that not only identifies enriched pathways from risk genes, but also uses these results to discover more genes.

CNVs collectively account for much larger genetic variation than SNVs and play an important role in neurodevelopmental disorders. Conceptually, CNVs and SNVs disrupting the same gene have similar effects and could be combined to map risk genes. In practice, though, interpretation of CNVs is not straightforward as CNVs often cover multiple genes. As result, CNV and SNV studies are almost always conducted separately. With exome and especially genome sequencing, our ability to detect CNVs has been dramatically increased. But without powerful tools to translate these data into direct knowledge of disease genes, we would waste a huge opportunity. **Aim 2 will provide, for the first time, a rigorous statistical framework to extract gene-level knowledge from CNVs and to link these findings with those from SNVs** for better identification of disease genes.

Expression QTL reveal the functional effects of genetic variants on gene expression, and is an important tool to extract mechanistic insights from GWAS. Aim 3 will develop a novel strategy for the joint analysis of eQTL and GWAS data, using only summary statistics. The main innovation is that **it exploits all eQTL of a gene, both *in cis* and *in trans*, to assess its role in a disease**. This gene-centered strategy circumvents the problem that GWAS findings often lead to no clear candidates genes. Furthermore, because most expression and phenotypic traits are highly polygenic, our ability to take advantage of weak loci, including those associated with gene expression *in trans*, is crucial to maximize the power of joint eQTL-GWAS analysis. Aim 3c will build an easy-to-use web based system: a user can simply upload a GWAS dataset, and the system would return all candidate genes and pathways by running our software on a large collection of eQTL datasets.

C. Approach

The research team: The PI developed a model for family sequencing data, that served as a main discovery tool in two large WES studies of autism [10, 16]. Dr. He has collaborated closely with autism experts, as attested by multiple publications [10, 46, 16, 47]. Dr. He has extensive experience in developing novel computational methods in statistical genetics [46, 39, 16], regulatory genomics [48, 49], evolutionary genomics [50, 51] and bioinformatics [52, 53]. The co-I, Dr. Matthew Stephens, has a long track record of creating innovative statistical methods in genetics, including the widely used tools for population structure analysis (Structure) and haplotype inference (PHASE) (cited more than 17,000 and 6,000 times, respectively) [54, 55].

Software development and distribution: We are committed to developing robust and user-friendly software for wide usage. The software will be developed in R, integrated with C++ for computationally intensive procedures. It will be distributed in user-friendly R packages. The PI will follow well-established software engineering principles (Dr. He has a PhD in Computer Science) to develop modular software that is easy to use and highly extendable.

Timeline: Work on all aims will start immediately after funding is available. In years 1-2, we will focus on method development. In year 3, we will focus on applying these methods in collaborative projects, refining the methods, and software finalization, documentation, testing and distribution.

Aim 1: Develop statistical methods for analyzing sequencing data from parent-child trios.

We start with a brief review of existing methods for analyzing *de novo* mutations from WES studies. The simplest strategy is to identify genes with recurrent (two or more) mutations [56, 57, 17]. This method will not scale well with large sample sizes (because more recurrent genes occur by chance), and does not account for gene-specific mutation rates. These problems can be remedied by a Poisson test, which compares the observed number of mutations in N families against the expected, $2N\mu$, where μ is the gene-specific mutation rate [46, 58]. In practice, researchers often focus on LoF mutations to enhance the signal.

Our overall strategy is to utilize all forms of genetic variation perturbing a gene. These would include LoF, missense and non-coding variants, both *de novo* and inherited from parents. The challenge is that these variants have very different effect sizes: e.g, a *de novo* LoF mutation in an autism gene on average increases the risk by 20 fold, while the same number for inherited missense variants is less than 2 [10]. Simply adding them in the

fashion of a Poisson test greatly dilutes the signal and reduces power. We recently developed TADA, the main analytic method empowering two large autism WES studies [10, 16]. TADA estimates average relative risks of different variant types, and uses the estimates to weigh them in a gene-level test. Building on the success of TADA, we will develop a more powerful framework with several major innovations: the ability to incorporate many variant annotations, a "spatial model" taking advantage of the clustering of risk variants in the genome, and a novel strategy for pathway analysis.

Preliminary studies

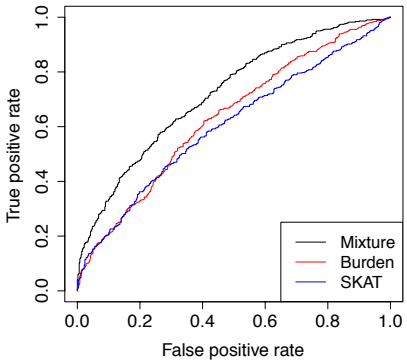

Figure 1: Classification of causal and non-causal genes by simulation (1000 cases and 1000 controls, $\gamma = 4$, $\pi = 0.2$).

Transmission and *De Novo* Association Test (TADA): TADA is a gene-level test using trio-sequencing data, combining information from d*de novo* mutations and inherited variants [46]. We start with the *de novo* model of TADA. Each mutation of a gene is assigned into a functional class, e.g. LoF, or damaging missense mutation. The number of mutations in a class c, denoted as x_c, follows: $x_c \sim \text{Poisson}(2N\mu_c\gamma_c)$, where N is sample size, μ_c the mutation rate of class c of the test gene, and γ_c the relative risk (RR) of class c. We assume that mutation rates are known [17, 58]. Under H_0 (non-risk gene), $\gamma_c = 1$, and under H_1 (risk gene), typically $\gamma_c > 1$. TADA estimates an average γ_c for each class: more deleterious classes on average have higher γ_c and hence larger contribution. TADA then computes a Bayes factor (BF) for each class c, comparing alternative against null models, and multiply these BFs to obtain the gene-level BF. TADA also has a likelihood model for inherited data, considering the number of inherited variants in each class, and similarly combines information from all classes. The total BF of a gene is the product of BFs from *de novo* and inherited models (see [46] for details).

We demonstrate through simulations that TADA almost doubles the power of identifying risk genes over the Poisson test, measured by the expected number of genes found at FDR < 0.1 when applying the methods genomewide. In a large WES study, we identified more than 100 putative risk genes at FDR < 0.3, most of which are new, enabling us to perform various pathway and network analyses [10].

Mixture model of variants: TADA does not model individual variants; instead, all variants in a class will be collapsed and treated as if they were a single variant. This is similar to burden test, commonly used in testing association of rare variants in a gene with phenotype [59, 60]. Using a simple mixture model, we show that this leads to loss of power, when risk variants are sparse. We assume a causal gene contains a mixture of risk and non-risk variants. For risk variants, their frequencies in cases are $\gamma > 1$ times higher than controls; for non-risk variants, their frequencies are equal. Non-causal genes in contrast have no risk variants. We assess the performance of this mixture model in classifying causal and non-causal genes using simulations, comparing it with burden test and SKAT [61]. When the ratio of risk variants, π, is small, the mixture model significantly outperforms burden test and SKAT (Figure 1). These results demonstrate the problem of variant collapsing, and highlight the opportunities of further improving TADA.

Burden analysis of non-coding *de novo* mutations in autism: We compare the rates of potential regulatory mutations in about 300 autistic children [62, 63, 19, 20] with 700 controls [64]. We use H3K27ac in developing brains to define enhancers [65] and other features, such as transcription factor motifs [66], to annotate mutations (SNVs). We found a general trend of enrichment of mutations in promoters ($<$ 1kb of TSS) and enhancers (within 10kb), with the results of motif-changing SNVs and of enhancers near constrained and ASD genes being statistically significant (Figure 2). These results show the importance of regulatory mutations in autism. The fact that the burdens vary by distance to TSS (higher burden in promoters),

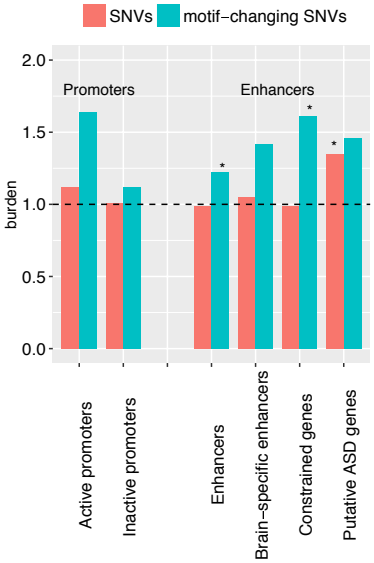

Figure 2: Burden analysis of de novo non-coding mutations. Active promoters are defined by H3K27ac marks while inactive promoters lack them (negative control). Constrained genes: depletion of genetic variations in population. Asterisks: $p < 0.05$.

effects on motifs (higher burden in mutations disrupting motifs) and tissue specificity of enhancers (higher burden in enhancers active only in brain), highlights the utilities of annotations to enrich causative mutations.

Relevant expertise of investigators: Besides the PI's direct experience with trio-sequencing studies, both PI and co-I have extensive experience in the statistical and genomics methods that we propose to exploit here. Our proposed model will involve Bayesian hierarchical model as a main statistical framework. The basic idea is that we can borrow information across variants and genes to learn important parameters. The co-I Dr. Stephens pioneered many applications of hierarchical models in statistical genetics [67, 68, 69, 70]. Both investigators are experts in regulatory genomics, which is important for interpretation of non-coding variants. Dr. He's thesis work on how regulatory elements control gene expression [48] forms the foundation of a research grant awarded to Saurabh Sinha by NIH (Sinha letter). Dr. He also published multiple papers on comparative genomics of regulatory elements [50, 71, 51]. Dr. Stephens is an investigator of the GTEx project, and has published influential papers on mapping human eQTL and chromatin accessibility QTL [67, 69, 72].

Aim 1a. Modeling individual variants while incorporating variant annotations

TADA collapses all variants in a class, ignoring heterogeneity in effects. As discussed previously, this could be detrimental to statistical power (Figure 1), a problem well recognized in rare variant association testing (RVAT) [73, 74, 60]. The problem is likely even worse for non-coding variants, whose effects are harder to predict and probably more heterogeneous. In principle this could be addressed by defining refined categories, e.g. "variants in promoters with low allele frequencies", but the number of categories grows exponentially with the number of annotations, so this approach is not scalable. Similar to the random effects models that are popular in RVAT [74, 61], we propose to generalize the TADA model to allow each variant to have a different effect. Specifically, each variant has a prior probability of being a causal variant, and this prior depends on annotations of variants.

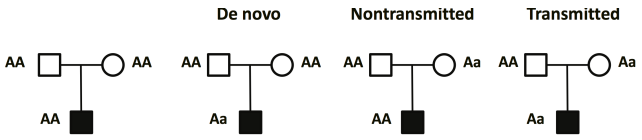

Figure 3: Four scenariors for a family trio with an affected child, where a is the minor (rare) allele.

Given a parent-child trio, there are four possible genotype combinations for a variant (Figure 3) – we consider only rare variants, thus homozygote mutants are very rare and ignored. Our data consists of the *de novo*, transmitted and non-transmitted counts for each variant, and variant annotations. For a test gene, let $x_j^{(d)}, x_j^{(1)}, x_j^{(0)}$ be these three counts for variant j. We define Z_j as the indicator variable of whether j is causal (1 if yes, 0 otherwise). The prior probability of Z_j, $\pi_j = P(Z_j = 1)$, is given by logistic regression: $\log(\pi_j/(1-\pi_j)) = \beta_0 + \sum_k \beta_k u_{jk}$, where u_{jk} is the k-th annotation of variant j and β_k is the effect of the k-th annotation. For a causal variant ($Z_j = 1$):

$$x_j^{(d)}|Z_j = 1 \sim \text{Poisson}(2N\mu_j\gamma_j) \quad x_j^{(1)}|(x_j^{(1)} + x_j^{(0)}), Z_j = 1 \sim \text{Binomial}\left(x_j^{(1)} + x_j^{(0)}, \gamma_j/(\gamma_j + 1)\right), \qquad (1)$$

where μ_j is mutation rate (known) and γ_j the relative risk (RR). The model of transmitted variants captures the intuition that risk variants are, on average, more likely to be transmitted from parents to offspring, the idea behind the Transmission Disequilibrium Test [75]. We assume the prior distribution: $\gamma_j \sim \text{Gamma}(\bar{\gamma}\sigma, \sigma)$, where $\bar{\gamma} > 1$ is the average RR. Under $Z_j = 0$, we simply have $\gamma_j = 1$ in the equation.

To identify risk genes, we compare H_0 of non-risk gene where $Z_j = 0$ for all variants vs. H_1 of risk gene, the model defined above assuming independence of variants. The result can be summarized as a Bayes Factor (BF) for the gene. To adjust for multiple testing, we use Bayesian false-discovery rate (FDR) control [76], and use simulations under the null to assess whether FDR is properly controlled. To estimate the parameters ($\bar{\gamma}$ and σ for the prior distribution of RRs and β_k for variant annotations), we propose to use Empirical Bayes, combining all variants across all genes. We will allow two different sets of parameters for *de novo* and inherited variants. To combine information across genes we use a mixture model: each gene is either causal or not, and the likelihood under each model is as given above.

Variant annotations: For coding variants, we consider: (1) predicted damaging effects from tools such as SIFT [77] and PolyPhen [78]; (2) tissue-specific expression of exons containing variants – e.g. a brain-expressed exon is likely relevant to psychiatric diseases [79]; (3) allele frequencies in human population [80]. For non-coding variants, we have two levels of annotations: for regulatory elements (usually several hundred bp long) and for individual nucleotides. The annotations of a variant would be the union of all annotations. The element-level

annotations include histone marks [81], chromatin accessibility [81], transcription factor binding [82], and so on. The nucleotide-level annotations include motif disruption and gains [83], evolutionary constraint [84, 85], allele frequencies, expression QTL [86], and aggregate scores such as C-scores [87] and FitCons [88].

Potential problems: One potential problem is that the annotations we use may not be sufficiently informative for distinguishing causal variants. Our strategies are: (1) We will explore the use of non-linear models such as Bayesian Adaptive Regression Trees to better combine annotations [89]. (2) Instead of estimating parameters using all genes, we can use high-confidence genes collected from prior studies to enhance statistical signal. (3) With a large number of annotations, we will explore the use of penalized likelihood to learn a "sparse" model (relatively few non-zero parameters), to avoid overfitting [90].

Aim 1b. Modeling spatial clustering of risk variants

Disease susceptibility variants often fall into spatial clusters, corresponding to natural units such as domains or exons for protein-coding sequences, and enhancers for non-coding sequences [91]. Leveraging spatial clustering patterns could boost power to map disease genes, and also help identify functional units that are particularly relevant to diseases of interest. We illustrate this with *de novo* mutations in non-coding regions, but the idea is similar for coding and inherited variants. We first identify putative functional elements around a gene, using a loose definition of "functional" (e.g. all regions in open chromatin). The remaining sequence will be ignored. A risk gene may have a large number of *de novo* mutations in all elements and this can be detected via our previous model (Figure 4A). Our new model will capture another scenario, where the total number of regulatory mutations is not significantly higher than chance expectation, but they are located within a single element (Figure 4B).

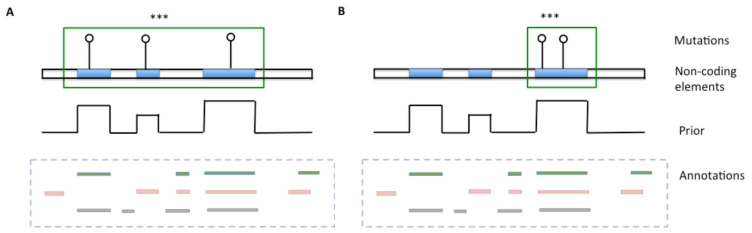

Figure 4: Non-coding model. Potential regulatory elements are shown in blue boxes. For simplicity, we assume all annotations are binary. (A) The example gene shows an enrichment of mutations in all regulatory elements. (B) Two mutations are located inside a single element, supporting the element being causal.

We propose to implement a two-level model to accommodate spatial clustering. The first level of the model is defined on elements: each element is either causal or not, denoted as $W_k \in \{0, 1\}$ for the k-th element, with the prior probability of being causal dependent on element-level annotations (Figure 4). That is, letting R_k denote element-level annotations for element k, we have:

$$P(x) = \prod_k \sum_{W_k} P(x_k|W_k)P(W_k), \quad \text{logit}P(W_k = 1) = R_k\theta, \tag{2}$$

where x denotes all data, θ is the effects of element-level annotations and x_k the data of element k. The second level of the model is defined on individual nucleotides and is essentially the model defined in Aim 1a and the details are skipped here. All the parameters of the model will be estimated using Empirical Bayes.

From this model, we can not only obtain the evidence supporting a gene, but also posterior probabilities for regulatory elements being causal (W_k). The knowledge of regulatory elements may provide additional insights of the disease mechanism: e.g. spatial-temporal activity patterns of these elements may suggest brain regions and developmental periods important for diseases.

Aim 1c. Incorporating gene-level annotations

To further improve identification of risk genes, we will develop a statistical procedure leveraging external annotations of genes that are informative of their roles in diseases such as pathways, tissue-specific expression, and known phenotypes. The input of this procedure is the BFs of all genes (from the analysis in previous steps) and their annotations. For the i-th gene, let Z_i be the indicator variable of whether it is a risk gene. Its prior, $\pi_i = P(Z_i = 1)$, depends on the annotations of the gene, A_i, through a logistic model: $\text{logit}(\pi_i) = A_i\theta$, where θ is

the parameters for annotations. Let B_i be the BF of gene i and x_i the data used to derive B_i. The likelihood is:

$$L(\theta) = \prod_i P(x_i|\theta) = \prod_i [\pi_i(\theta)P(x_i|Z_i = 1) + (1 - \pi_i(\theta))P(x_i|Z_i = 0)] \quad (3)$$

assuming independence. We factorize the term $\prod_i P(x_i|Z_i = 0)$ (which does not depend on θ), and have: $L(\theta) \propto \prod_i [\pi_i(\theta)B_i + (1 - \pi_i(\theta))]$. This likelihood can be maximized using an EM algorithm. The estimated parameters $\hat{\theta}$ reveal the most relevant gene annotations (e.g. a pathway related to disease). We can further obtain the posterior of each gene by Bayes Theorem: $P(Z_i = 1|x_i) \propto \pi_i(\hat{\theta})B_i$. Thus our final statistical measure of a gene combines information from the sequencing data (BF) and prior annotations.

The advantage of our procedure over the standard "enrichment or pathway analysis" [92] is that it can use the results of enrichment to re-rank genes: genes with enriched annotations will have higher prior probabilities. Furthermore, it is more sensitive in that it avoids the use of a hard cutoff to define "significant" genes. The benefits of this overall strategy have been demonstrated by Dr. Stephens in the context of GWAS [68].

Validation and evaluation

We will use simulations to assess how the power of our methods depends on factors such as the fraction of risk variants and the extent of spatial clustering. We will also investigate the robustness to violations of assumptions: e.g. a different prior distribution of variant relative risks. We will compare our methods with TADA, Poisson test, and a recently developed method, FitDNM [93] (which is limited to *de novo* coding mutations, though).

We will apply the methods to autism WES and WGS data from Autism Sequencing Consortium and Simons collection (Buxbaum letter). We will use publicly available annotations such as ENCODE [82] and PsychENCODE [94], as well as regulatory information most relevant to autism from collaborators (Noonan letter). Our results will be compared with existing methods above, using the number of discovered genes at an FDR as the metric. We will evaluate predicted genes in multiple ways: (1) their roles in related disorders such as schizophrenia; (2) their expression patterns in brain using BrainSpan [79]; (3) evolutionary constraint, which is a strong marker of autism genes; (4) their relations with known autism genes, using network analysis tools such as GeneMania [95].

The model also identifies regulatory elements, and we will evaluate them in multiple ways, e.g. by assessing their overlap with GWAS loci of psychiatric diseases [4, 3], and with brain eQTL [96, 97, 98]. We will also pursue experimental characterization, through collaborations, on promising regulatory elements (Noonan letter).

Aim 2: Develop methods that leverage copy number variations (CNVs) for gene mapping

Unlike SNVs, the length of CNVs vary enormously, from 50 bp to tens of millions bps. This poses unique and important analytic challenges. As a result, there is no single commonly accepted method for disease mapping of CNVs. For case-control studies, one can associate the status of a CNV or a genomic region (presence or absence) with the disease status [99, 21]. For *de novo* CNVs, typically researchers declare a CNV "pathogenic" if it is large (> 500 kb) and rare in the population [21], but it is often difficult to firmly link a *de novo* CNV with a specific disorder. In all these cases, the results are candidate CNVs or regions, not genes.

Another strategy is to test the "CNV burden" of a gene set: whether the genes are disrupted by CNVs more often in cases than in controls [22, 25, 100]. However, this analysis is error-prone. For example, case subjects tend to have more and bigger CNVs than controls, thus a naive burden analysis may find many gene sets due to overall higher burden in cases (see [25] for discussion of these problems). Adopting such tests to single genes would be even more problematic as adjacent genes tend to be targeted by the same CNVs. Thus several neighboring genes may all have significant burdens; but the signal is explained by a single risk gene. Combining SNV and CNV data creates additional challenges, because the effect sizes of SNVs and CNVs targeting a gene are generally very different. Simply adding the number of SNVs and CNVs would lead to loss of power. While gene-level analysis from CNVs were attempted [99, 101], the underlying statistical issues remain unaddressed.

Overview of our approach. Given CNVs from *de novo* or case-control studies, our goal is to infer the posterior probabilities of genes affecting the disease risk. These posteriors can be easily integrated with another gene-level study: e.g. we can treat the posteriors of genes as prior probabilities in the second study.

Our proposed procedure conceptually has two steps. First, we will do CNV-level analyses to estimate the evidence for each CNV being disease-related. Instead of calling a CNV causal or not, we will express evidence

as probabilities. Next, we find the best configurations of genes (the vector of variables indicating whether each gene is a risk gene) consistent with the evidence for CNVs. This is based on the logic that a disease-related CNV must contain at least one risk gene. For example, suppose we find two causal CNVs in a region, overlapping genes A, B and B, C, respectively, we can infer that B is the risk gene driving the evidence of both CNVs. The best configurations need not be unique, and our model infers the most probable configurations.

Preliminary studies

In a recent study, we found that the genes with *de novo* LoF mutations are four-fold enriched with *de novo* deletions in autistic children, suggesting that *de novo* SNVs and CNVs often target the same genes [16]. To exploit this pattern, we developed a procedure to combine CNV and SNV data at the gene level: (1) For each CNV, we assess its likelihood of contributing risk. We use the TADA model, effectively treating a CNV as a "gene", and estimate its BF. (2) CNVs often overlap multiple genes, so we developed a heuristic formula to "discount" the BF of a CNV when evaluating its member genes, based on its size (larger CNVs carry weaker evidence of individual genes and will be discounted more). The result is a BF for each member gene of any CNV. (3) For a gene of interest, we combine the evidence of all CNVs containing it by multiplying the BFs of this gene from step 2. The resulting BF is then multiplied to the BF of the same gene from SNV data to obtain the final gene-level BF.

We applied this approach to an assembled WES and CNV dataset of autistic trios. Comparing with using SNV data alone, we found six more genes, most representing highly promising candidate genes [16].

Detailed approach

The method outlined above was designed for *de novo* CNVs and cannot cope with case-control studies. Even with *de novo* CNVs, the method has significant limitations. It assumes one causal gene per CNV, which is not valid for large CNVs. It analyzes one CNV a time, which is equivalent to the assumption of statistical independence of CNVs. With overlapping CNVs, this assumption is violated, and the results could be invalid BFs of genes.

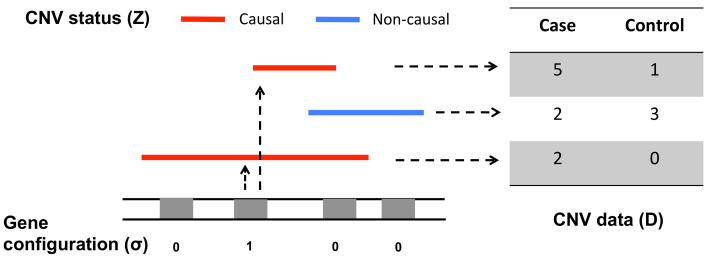

Figure 5: CNV model, showing a region of 4 genes and three CNVs. The vertical arrows show that the CNV status are determined by the gene configuration, and the horizontal arrows $P(D|Z)$. We use a compact representation of CNV data: only the counts, instead of individual-level CNVs and phenotypes.

For simplicity, suppose the genome is divided into disjoint regions (i.e. no CNVs in common between regions). Within a region R, we may have multiple, possibly overlapping, CNVs. We use a generic symbol D for our CNV data. In the *de novo* case, D means the number of *de novo* events, d_j, for the j-th CNV. In the case-control case, D represents CNVs and phenotypic data: x_{ij} for the presence or absence of j-th CNV in sample i, and y_i the phenotypic status of i. We define σ as a binary vector indicating whether each gene in R is a risk gene or not, called "configuration". Our goal is to infer the posterior distribution $P(\sigma|D)$.

The intuition of our model is: gene configuration determines the susceptibility status of CNVs (CNV configuration), denoted as $Z_j \in \{0, 1\}$ for the j-th CNV, which in turn influence the pattern of CNV data (Figure 5). We can write this as $\sigma \to Z \to D$. The distribution $P(Z|\sigma)$ expresses the consistency of σ and Z: a CNV is causal if and only if at least one of its member gene(s) is a risk gene, or: $P(Z|\sigma) = 1$ if Z and σ are consistent and 0 otherwise. The distribution $P(D|Z)$ is specified below. For simplicity, we define Z_0 as the special CNV configuration where $Z_j = 0$ for all CNVs, and write $P(D|Z) = P(D|Z_0)B(Z)$, where $B(Z)$ is similar to Bayes factor, expressing the support of the CNV configuration Z. Now we can infer the posterior of σ given D using Bayes Theorem. We denote Z_σ as the CNV configuration consistent with σ (unique once σ is given):

$$P(\sigma|D) \propto P(\sigma)P(D|\sigma) = P(\sigma)\sum_Z P(D|Z)P(Z|\sigma) = P(\sigma)P(D|Z_\sigma) \propto P(\sigma)B(Z_\sigma). \quad (4)$$

The prior $P(\sigma)$ is independent Bernoulli distributions: each gene has a certain prior probability of being a risk gene. Our inference has two steps: first, obtain the evidence of all CNVs in a region, $B(Z)$, from CNV data; and then use the equation above to translate $B(Z)$ into posteriors of σ. Below we describe details of the first step.

Case-control model: To infer CNV configuration $B(Z)$ from case-control data, we leverage the statistical machinery of Bayesian regression. Specifically, let β_j be the effect size of the j-th CNV. We assume a spike-and-slab prior for β_j and logistic regression model for the phenotype:

$$\beta_j|Z_j = 0 \sim 0, \qquad \beta_j|Z_j = 1 \sim N(0, \sigma^2), \qquad \text{logit}P(y_i = 1) = \beta_0 + \rho w_i + \sum_j \beta_j x_{ij}, \tag{5}$$

where w_i denotes possible covariates (such as the total number of CNVs in subject i) and ρ their effects. Inference on this model leads to the support of each CNV configuration $B(Z)$.

***De novo* model**: We use the *de novo* mutation model of TADA. Specifically, let γ_j be the relative risk of CNV j, if $Z_j = 0$, $\gamma_j = 1$ and if $Z_j = 1$, γ_j follows a Gamma prior distribution. The model of *de novo* CNV count is $d_j \sim$ Poisson$(2N\mu_j\gamma_j)$ where N is sample size and μ_j the rate of CNV event.

Potential problems: We assume the parameters of CNV effects, β_j or γ_j above, are the same across all CNVs. In reality, the effect of a CNV may depend on its size and whether it is deletion or insertion. We will explore more flexible models that assign different priors for different kinds of CNVs. Another challenge is that there are 2^K possible configurations for a region of K genes, and computation of $P(\sigma|D)$ could be prohibitive. An approximation strategy is to enumerate only configurations where each region has only a small number of causal genes. Alternatively, we will use Monte Carlo strategy to sample from the posterior distribution $P(\sigma|D)$.

Validation and evaluation

We will first use simulations to assess our method. We start by sampling risk genes from the genome, then sample the locations of CNVs to match the sizes and frequencies of CNVs in real data. Next we will sample the effect sizes of CNVs based on whether they contain risk genes and generate CNV data. We will evaluate if the BFs of non-risk genes are inflated [102], and compare the power of our method with related methods [25, 100].

We will apply the method to several CNV datasets, including those from Simons collection [16] (Sanders letter) and PGC (Sullivan letter). In both cases, we will combine CNV data with SNV data from WES or WGS (Aim 1) to assess the benefit of joint SNV-CNV analysis. The results will be validated in two ways: first, we can infer the posterior probabilities of CNVs $P(Z|D)$, and compare results with known pathogenic CNVs. Second, we will evaluate the predicted autism or schizophrenia risk genes using strategies outlined in Aim 1.

Aim 3. Develop methods for integrated analysis of GWAS and expression QTL

GWAS, even when successful, seldom reveal causal variants and gene targets. Our goal here is to use independent eQTL data to re-interpret GWAS to identify risk genes. To make the method widely applicable, we use summary statistics in the form of effect sizes or p-values. This is important as summary statistics are much easier to share than individual level data. Because expression and phenotypes are often not collected in the same individuals, we cannot directly test their relationship [103]. Instead, our inference must rely on the pattern of shared association: SNPs showing signs of association with both expression and phenotype. The challenge is that three scenarios could lead to this pattern (Figure 6A): causality of gene, pleiotropic effects, or LD between two causal variants. Another important consideration is that a gene may have multiple eQTL. In particular, *trans*-eQTL are individually weaker than *cis*-eQTL, but collectively account for a larger proportion of expression variation [35, 36]. Our proposed method models the three scenarios, and aggregates information across all eQTL of a gene.

A few methods have been published to integrate eQTL and GWAS data at the gene level: (1) Gene expression imputation, such as PrediXcan [32] and TWAS [33]. The ideas behind these two methods are similar: use the eQTL data to train a predictive model of gene expression from genotypes; apply this model in the GWAS data to "impute" gene expression level; and then test correlation of imputed expression and phenotype in the GWAS cohort. These methods uncover only correlation, but not causality, between gene and phenotype through shared loci, and are currently limited to only *cis*-eQTL. (2) Mendelian randomization. The SMR method uses the top *cis*-eQTL of a gene as the "instrumental variable" to assess the effect of the gene on the phenotype [34]. It uses information only from one locus of a gene, and cannot distinguish causality from pleiotropy.

Overview of our approach. If we consider only one shared locus between gene and phenotype, causality and pleiotropy are virtually indistinguishable, as others pointed out [33]. A defining feature of our proposed approach is that we simultaneously model all eQTL of a gene: both *cis*- and *trans*-eQTL of a wide range of effect sizes.

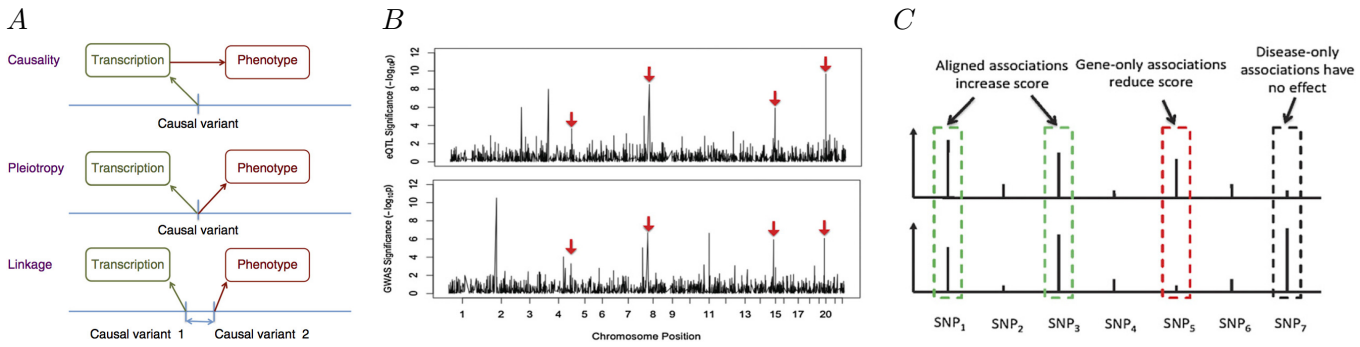

Figure 6: Integrating summary statistics of eQTL and GWAS. (A) Three possible scenarios at one locus associated with expression and pheontype. Adapted from [34]. (B) The eQTL of a causal gene overlap with the GWAS loci of the phenotype (hypothetical example). (C) Alignment of the association statistics of an expression trait (top) and the phenotype (bottom).

Consider a causal gene of a disease with multiple eQTL. Conceptually, change of genotype at any of these eQTL will lead to change of expression level, which may in turn alter the disease risk. Therefore, many of these eQTL are likely to be also associated with the disease. Furthermore, the effect sizes of a SNP with respect to gene expression and phenotype should be correlated: a stronger eQTL should lead to a larger effect on the pheotype. Our inference of causal genes is based on this signature of matched SNPs and correlated effects across multiple loci in the genome (Figure 6B). Non-causal genes, in contrast, might share some SNPs with the phenotype, but lack this overall pattern of correlation. The success of this strategy depends on a relatively large number of independent eQTL. So we model not only eQTL passing threshold, but also weaker ones, taking advantage of polgenicity of expression traits [42]. **In summary, our approach allows researchers to analyze GWAS from a novel angle by taking advantage of eQTL, opening up a new direction of disease gene discovery.**

Preliminary studies

Sherlock. We previously developed a method, Sherlock, for combining eQTL and GWAS data [39]. It takes as input the p-values from independent eQTL and GWAS data, one gene a time. Intuitively, Sherlock performs an "alignment" of two sets of p-values. For any SNP, there are three possible scenarios (Figure 6C). A SNP associated with both traits contributes a positive score to the gene. A SNP associated with only one trait either contributes a negative score (counter-evidence against the gene) or has no contribution (not informative). The total score of a gene is the sum of scores from all eQTL of this gene (we use a loose p-value cutoff for defining eQTL). Mathematically, Sherlock implements this idea using a probabilistic graphical model and expresses the evidence of genes as Bayes factors.

Table 1: Supporting SNPs of the gene EFS in Sherlock analysis of Crohn's Disease. p-eQTL: p-values of association with EFS expression.

Position	cis/trans	p-eQTL	p-GWAS
chr4:32433817	trans	9×10^{-6}	8.8×10^{-4}
chr14:33091890	trans	1×10^{-5}	3.3×10^{-5}
chr16:75957583	trans	1×10^{-6}	9.6×10^{-4}
chr21:29837833	trans	6×10^{-6}	1.2×10^{-4}

We applied Sherlock to GWAS of Crohn's disease and type 2 diabetes, using lymphoblast and liver eQTL, respectively. In each analysis, we identified new disease genes that were missed by GWAS. One striking example is the gene EFS, which has no association signal in GWAS of Crohn's disease. All supporting SNPs of EFS are located in trans and the p-values of these SNPs in both eQTL and GWAS are modest by the GWAS standard (Table 1). EFS knockout mice develop inflammatory lesions in small intestine, a pattern very similar to Crohn disease [104, 105]. Another example is LYNX1, which was also found entirely through *trans*-eSNPs and is being explored as a therapeutic target for treating Crohn's disease [106].

Regression with Summary Statistics (RSS). Dr. Stephens' group recently developed a statistical framework, RSS, for inference on summary statistics, in the form of effect sizes from univariate regression analysis (i.e. testing one SNP a time). Its key component is a likelihood model relating the true effect sizes and the observed univariate effects, accounting for correlation of summary statistics of adjacent SNPs due to linkage disequilibrium. We have implemented methods for fine-mapping and estimation of heritability [107] and pathway analysis (unpublished data). In all cases, the results are similar to those from analysis with individual-level data.

Methods for eQTL mapping. Having high-quality eQTL is essential for the success of our approach. Dr. Stephens is a member of the Analysis Working group, and the Steering Committee of the GTEx project [29]. Dr. Stephens has developed powerful methods for eQTL mapping jointly in multiple tissues, borrowing information across tissues [69]. His group has also been involved in developing methods that integrate functional annotations of variants, such as chromatin accessibility, to increase the power of eQTL mapping [67].

Aim 3a. Integrated GWAS-eQTL analysis for individual genes

Despite its potential, Sherlock has weaknesses that limit its application. First, Sherlock has a very simple model of LD and cannot distinguish between linkage (Figure 6A, bottom) and true co-localization of association signals. Second, Sherlock works with p-values, when effect sizes would be more informative. One expects that for eQTL of a risk gene, the magnitude of effects on expression should be correlated with that on phenotype. Finally, in Sherlock, important parameters (Bayesian hyperprior parameters) are fixed instead of being estimated from data, and as a result, the Bayes factors of genes are not always calibrated. The goal of this sub-aim is to address these limitations, leading to a method that is more powerful and widely applicable.

Model of independent loci: For simplicity, we assume all SNPs are independent. We have summary statistics of eQTL of a gene: $\hat{\theta}_i$ for the observed effect size of SNP i from univariate analysis, $1 \leq i \leq n$, and s_i its standard error. Similarly, we have $\hat{\beta}_i$ and t_i for the effect size of SNP i in GWAS and its standard error. The true effects in eQTL and GWAS (unobserved) are denoted as θ_i and β_i, respectively. The observed univariate effects are related to the true effects by: $\hat{\theta}_i|\theta_i \sim N(\theta_i, s_i^2), \hat{\beta}_i|\beta_i \sim N(\beta_i, t_i^2)$. Our key idea is that for non-causal genes, θ_i and β_i are independent; but for causal genes, they are correlated. We propose to use a Bayesian sparse prior (spike-and-slab) for true eQTL effects, and a linear model for GWAS effects:

$$\theta_i \sim (1-\pi_\theta)\delta_0 + \pi_\theta N(0, \sigma_\theta^2) \qquad \beta_i|\theta_i \sim N(\lambda\theta_i, \tau^2), \tag{6}$$

where π_θ is the fraction of causal eQTL (which may vary among genes) and δ_0 Dirac's delta function at 0. The parameter λ can be interpreted as the effect of gene expression change on the phenotype and τ^2 the deviation of phenotypic effect from expectation. To understand our model, consider an eQTL Q_i, gene expression E and phenotype P with this model: $Q_i \xrightarrow{\theta_i} E \xrightarrow{\lambda} P$, where θ_i and λ are effect sizes. Assuming simple linearity and ignoring error terms, we have: $P = \lambda E = \lambda(\theta_i Q_i) = (\lambda\theta_i)Q_i$. So the effect of Q_i on P is $\lambda\theta_i$ as appears above.

For our model of non-causal genes, we replace the prior of β_i in Equation 6 with a sparse prior similar to eQTL effects: $\beta_i \sim (1-\pi_\beta)\delta_0 + \pi_\beta N(0, \sigma_\beta^2)$. To test if a gene is causal, we compute the conditional probability, $P(\hat{\beta}|\hat{\theta}, \mathbf{s}, \mathbf{t})$ under causal or non-causal model, where $\hat{\beta}$ and $\hat{\theta}$ are the vectors of observed summary statistics, and $\mathbf{s}, \mathbf{t}$ the vectors of standard errors. This involves integration of θ_i and β_i, which can be done analytically (not shown). The Bayesian hyperprior parameters in eQTL and GWAS, $\pi_\theta, \pi_\beta, \sigma_\theta^2, \sigma_\beta^2$ can be estimated from summary statistics using RSS. The parameters λ and τ^2 can be estimated from all the SNPs of a gene.

Model in the presence of LD: We use the RSS model to handle the situation where SNPs are in LD [107]. Specifically, our prior model for θ and β (vector of true eQTL and GWAS effects) are exactly the same as Equation 6. The likelihood functions $P(\hat{\theta}|\theta)$ and $P(\hat{\beta}|\beta)$ come from the RSS model, accounting for LD of the SNPs - see Zhu et al. [107]. Our inference is still based on $P(\hat{\beta}|\hat{\theta}, \mathbf{s}, \mathbf{t})$ under null and alternative models, marginalizing the true effects. This probability no longer has closed form. Instead, we can use the developed RSS method to obtain posterior sample of the locations of all true eQTL of a gene (the implicit binary indicators in Equation 6). Once these are given, one can show that the parameters β and θ can still be marginalized analytically.

Understanding the model: We offer some intuitive explanations of how our model distinguishes the three scenarios in Figure 6A. For a causal gene, most of its eQTL should be shared with GWAS, as explained previously. Any shared loci (i.e. large eQTL and GWAS effects) are consistent with our causal model, Equation 6, elevating the gene's evidence. A pleiotropic gene, on the other hand, might share some loci, but most of their eQTL have little effect in GWAS. This pattern of non-shared eQTL is against our expectation under Equation 6, and as a result, our model will penalize this gene. Our method further distinguishes "linkage" and"causality". Under the causality scenario, we expect true eQTL and GWAS effects (θ and β) are correlated; but under linkage, the true effects are independent and the colocalization of observed eQTL and GWAS effects ($\hat{\theta}$ and $\hat{\beta}$) are only due to LD.

Potential problems: The key assumption of our approach is the linear relation between expression and phenotypic effects, specified in Equation 6. We can imagine scenarios where this assumption may be violated. For

example, when considering disease risk (as opposed to quantitative trait), increasing or decreasing of expression of a causal gene may both increase the disease risk (making expression level sub-optimal). We will explore more flexible priors, for instance, for binary traits, considering only the magnitude of effects and ignoring directions.

Evaluation: We will assess, via simulation, how the power of our method depends on various factors such as the number of causal eQTL and GWAS loci, the SNP effect sizes, and the effect size at gene level (λ). We will also assess how robust the model is to violation of assumptions, e.g. the linear relation between the two effect sizes, as explained above. In all simulations, we will compare the power and false positive rates with related methods including COLOC [108], PrediXcan [32], TWAS [33] and SMR [34].

We will apply our method to GWAS data of schizophrenia from PGC (Sullivan letter), using several brain eQTL datasets (Liu letter). It was shown recently that schizophrenia may involve processes outside brain, especially the immune system, so we will also use eQTL from other tissues in GTEx [109]. We will evaluate our predictions in several ways. First, we will replicate results using an independent GWAS dataset (or cross-validation), and GWAS of related phenotypes (e.g. bipolar disorder from PGC). By replicate, we mean that a predicted gene has some statistical support in a different dataset, even if the supporting SNPs differ. Next, we will evaluate brain expression of predicted genes: e.g. if they are differentially expressed in patients [110]. We will also search for published results and other resources for evidence regarding the predicted genes.

Aim 3b. Integrated GWAS-eQTL analysis for pathways

This sub-aim tests if a group of genes (pathway), has an effect on a phenotype. Pathway analysis is a major tool in GWAS, as it increases power compared with analysis of individual SNPs and the results can be more interpretable [92]. In our case, pathway-level analysis is particularly interesting. Functionally related genes are often similarly regulated, thus may share similar eQTLs. In the extreme case, one can imagine a group of genes with nearly identical eQTL patterns, and it would be difficult to know which gene is a driver of the disease. In such scenarios, a more meaningful question is whether the group, as a whole, affects the disease risk.

We denote θ_{ij} as the (true) effect of SNP i on expression of gene j ($1 \leq j \leq m$), λ_j the effect of gene j on the phenotype, and β_i the GWAS effect of SNP i. The relationship of GWAS and eQTL effects is now modeled by multiple regression, a direct generalization of the simple linear model of the two described in Equation 6:

$$\theta_{ij} \sim (1-\pi_{\theta j})\delta_0 + \pi_{\theta j} N(0, \sigma^2_{\theta j}) \qquad \beta_i|\theta_{\mathbf{i}} \sim N\left(\sum_j \theta_{ij}\lambda_j, \tau^2\right), \tag{7}$$

where $\theta_{\mathbf{i}}$ is the effects of SNP i on m genes. Note that the fraction of causal variants and the effect size priors are gene-specific. Next, the observed eQTL and GWAS effects are related to true effects by the same models described in Aim 3a. Our problem then becomes testing the null hypothesis that $\lambda_j = 0$ for each gene j. This is based on $P(\hat{\beta}|\hat{\theta}, \lambda)$, again marginalizing the true effects.

Evaluation: We will first evaluate our method by simulations. We will use real eQTL data to preserve the pattern of eQTL similarity among genes and simulate GWAS data. We will study a few scenarios: e.g. each gene in a pathway has a small effect (λ_j), or a small number of genes in a pathway have large effects. The power and false positive rates will be evaluated. To apply the method to psychiatric diseases, we will obtain gene pathways from various public collections [111]. In particular, we will obtain gene clusters from co-expression analysis in brain [110]. The predicted pathways will be evaluated in multiple ways, similar to our single-gene analysis, such as replication in independent GWAS or GWAS of related phenotypes, and differential expression in patients.

Aim 3c. Building a web-based system for GWAS summary statistics analysis using eQTL

Many eQTL studies have been published, but there are few repositories for researchers to quickly access diverse eQTL datasets. For those that exist, user interaction is always based on variants: find genes associated with a variant, or variants associated with a gene [112, 113]. This sub-aim will build an online system that supports downstream gene-level analysis of GWAS using eQTL. We will curate publicly available eQTL datasets, and precompute summary statistics and important parameters. We will integrate these resources with our software and create a web interface. A user can upload GWAS summary statistics, and our system will return candidate genes or pathways, by comparing GWAS with each of our curated eQTL datasets.

References

[1] The NHGRI-EBI Catalog of published genome-wide association studies. https://www.ebi.ac.uk/gwas/.

[2] J. Gratten, N. R. Wray, M. C. Keller, and P. M. Visscher. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.*, 17(6):782–790, Jun 2014. PMCID: PMC4112149.

[3] S. H. Lee, S. Ripke, B. M. Neale, S. V. Faraone, S. M. Purcell, R. H. Perlis, B. J. Mowry, A. Thapar, M. E. Goddard, J. S. Witte, D. Absher, I. Agartz, H. Akil, F. Amin, O. A. Andreassen, A. Anjorin, R. Anney, V. Anttila, D. E. Arking, P. Asherson, M. H. Azevedo, L. Backlund, J. A. Badner, A. J. Bailey, T. Banaschewski, J. D. Barchas, M. R. Barnes, T. B. Barrett, N. Bass, A. Battaglia, M. Bauer, M. Bayes, F. Bellivier, S. E. Bergen, W. Berrettini, C. Betancur, T. Bettecken, J. Biederman, E. B. Binder, D. W. Black, D. H. Blackwood, C. S. Bloss, M. Boehnke, D. I. Boomsma, G. Breen, R. Breuer, R. Bruggeman, P. Cormican, N. G. Buccola, J. K. Buitelaar, W. E. Bunney, J. D. Buxbaum, W. F. Byerley, E. M. Byrne, S. Caesar, W. Cahn, R. M. Cantor, M. Casas, A. Chakravarti, K. Chambert, K. Choudhury, S. Cichon, C. R. Cloninger, D. A. Collier, E. H. Cook, H. Coon, B. Cormand, A. Corvin, W. H. Coryell, D. W. Craig, I. W. Craig, J. Crosbie, M. L. Cuccaro, D. Curtis, D. Czamara, S. Datta, G. Dawson, R. Day, E. J. De Geus, F. Degenhardt, S. Djurovic, G. J. Donohoe, A. E. Doyle, J. Duan, F. Dudbridge, E. Duketis, R. P. Ebstein, H. J. Edenberg, J. Elia, S. Ennis, B. Etain, A. Fanous, A. E. Farmer, I. N. Ferrier, M. Flickinger, E. Fombonne, T. Foroud, J. Frank, B. Franke, C. Fraser, R. Freedman, N. B. Freimer, C. M. Freitag, M. Friedl, L. Frisen, L. Gallagher, P. V. Gejman, L. Georgieva, E. S. Gershon, D. H. Geschwind, I. Giegling, M. Gill, S. D. Gordon, K. Gordon-Smith, E. K. Green, T. A. Greenwood, D. E. Grice, M. Gross, D. Grozeva, W. Guan, H. Gurling, L. De Haan, J. L. Haines, H. Hakonarson, J. Hallmayer, S. P. Hamilton, M. L. Hamshere, T. F. Hansen, A. M. Hartmann, M. Hautzinger, A. C. Heath, A. K. Henders, S. Herms, I. B. Hickie, M. Hipolito, S. Hoefels, P. A. Holmans, F. Holsboer, W. J. Hoogendijk, J. J. Hottenga, C. M. Hultman, V. Hus, A. Ingason, M. Ising, S. Jamain, E. G. Jones, I. Jones, L. Jones, J. Y. Tzeng, A. K. Kahler, R. S. Kahn, R. Kandaswamy, M. C. Keller, J. L. Kennedy, E. Kenny, L. Kent, Y. Kim, G. K. Kirov, S. M. Klauck, L. Klei, J. A. Knowles, M. A. Kohli, D. L. Koller, B. Konte, A. Korszun, L. Krabbendam, R. Krasucki, J. Kuntsi, P. Kwan, M. Landen, N. Langstrom, M. Lathrop, J. Lawrence, W. B. Lawson, M. Leboyer, D. H. Ledbetter, P. H. Lee, T. Lencz, K. P. Lesch, D. F. Levinson, C. M. Lewis, J. Li, P. Lichtenstein, J. A. Lieberman, D. Y. Lin, D. H. Linszen, C. Liu, F. W. Lohoff, S. K. Loo, C. Lord, J. K. Lowe, S. Lucae, D. J. MacIntyre, P. A. Madden, E. Maestrini, P. K. Magnusson, P. B. Mahon, W. Maier, A. K. Malhotra, S. M. Mane, C. L. Martin, N. G. Martin, M. Mattheisen, K. Matthews, M. Mattingsdal, S. A. McCarroll, K. A. McGhee, J. J. McGough, P. J. McGrath, P. McGuffin, M. G. McInnis, A. McIntosh, R. McKinney, A. W. McLean, F. J. McMahon, W. M. McMahon, A. McQuillin, H. Medeiros, S. E. Medland, S. Meier, I. Melle, F. Meng, J. Meyer, C. M. Middeldorp, L. Middleton, V. Milanova, A. Miranda, A. P. Monaco, G. W. Montgomery, J. L. Moran, D. Moreno-De-Luca, G. Morken, D. W. Morris, E. M. Morrow, V. Moskvina, P. Muglia, T. W. Muhleisen, W. J. Muir, B. Muller-Myhsok, M. Murtha, R. M. Myers, I. Myin-Germeys, M. C. Neale, S. F. Nelson, C. M. Nievergelt, I. Nikolov, V. Nimgaonkar, W. A. Nolen, M. M. Nothen, J. I. Nurnberger, E. A. Nwulia, D. R. Nyholt, C. O'Dushlaine, R. D. Oades, A. Olincy, G. Oliveira, L. Olsen, R. A. Ophoff, U. Osby, M. J. Owen, A. Palotie, J. R. Parr, A. D. Paterson, C. N. Pato, M. T. Pato, B. W. Penninx, M. L. Pergadia, M. A. Pericak-Vance, B. S. Pickard, J. Pimm, J. Piven, D. Posthuma, J. B. Potash, F. Poustka, P. Propping, V. Puri, D. J. Quested, E. M. Quinn, J. A. Ramos-Quiroga, H. B. Rasmussen, S. Raychaudhuri, K. Rehnstrom, A. Reif, M. Ribases, J. P. Rice, M. Rietschel, K. Roeder, H. Roeyers, L. Rossin, A. Rothenberger, G. Rouleau, D. Ruderfer, D. Rujescu, A. R. Sanders, S. J. Sanders, S. L. Santangelo, J. A. Sergeant, R. Schachar, M. Schalling, A. F. Schatzberg, W. A. Scheftner, G. D. Schellenberg, S. W. Scherer, N. J. Schork, T. G. Schulze, J. Schumacher, M. Schwarz, E. Scolnick, L. J. Scott, J. Shi, P. D. Shilling, S. I. Shyn, J. M. Silverman, S. L. Slager, S. L. Smalley, J. H. Smit, E. N. Smith, E. J. Sonuga-Barke, D. St Clair, M. State, M. Steffens, H. C. Steinhausen, J. S. Strauss, J. Strohmaier, T. S. Stroup, J. S. Sutcliffe, P. Szatmari, S. Szelinger, S. Thirumalai, R. C. Thompson, A. A. Todorov, F. Tozzi, J. Treutlein, M. Uhr, E. J. van den Oord, G. Van Grootheest, J. Van Os, A. M. Vicente, V. J. Vieland, J. B. Vincent, P. M. Visscher, C. A. Walsh, T. H. Wassink, S. J. Watson, M. M. Weissman, T. Werge, T. F. Wienker, E. M. Wijsman, G. Willemsen, N. Williams, A. J. Willsey, S. H. Witt, W. Xu, A. H. Young, T. W. Yu, S. Zammit, P. P. Zandi, P. Zhang, F. G. Zitman, S. Zollner, B. Devlin, J. R. Kelsoe, P. Sklar, M. J. Daly, M. C. O'Donovan, N. Craddock, P. F. Sullivan, J. W. Smoller, K. S. Kendler, and N. R. Wray. Genetic relationship between five

psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.*, 45(9):984–994, Sep 2013. PMCID: PMC3800159.

[4] S. Ripke, B. M. Neale, A. Corvin, J. T. Walters, K. H. Farh, P. A. Holmans, P. Lee, B. Bulik-Sullivan, D. A. Collier, H. Huang, T. H. Pers, I. Agartz, E. Agerbo, M. Albus, M. Alexander, F. Amin, S. A. Bacanu, M. Begemann, R. A. Belliveau, J. Bene, S. E. Bergen, E. Bevilacqua, T. B. Bigdeli, D. W. Black, R. Bruggeman, N. G. Buccola, R. L. Buckner, W. Byerley, W. Cahn, G. Cai, D. Campion, R. M. Cantor, V. J. Carr, N. Carrera, S. V. Catts, K. D. Chambert, R. C. Chan, R. Y. Chen, E. Y. Chen, W. Cheng, E. F. Cheung, S. A. Chong, C. R. Cloninger, D. Cohen, N. Cohen, P. Cormican, N. Craddock, J. J. Crowley, D. Curtis, M. Davidson, K. L. Davis, F. Degenhardt, J. Del Favero, D. Demontis, D. Dikeos, T. Dinan, S. Djurovic, G. Donohoe, E. Drapeau, J. Duan, F. Dudbridge, N. Durmishi, P. Eichhammer, J. Eriksson, V. Escott-Price, L. Essioux, A. H. Fanous, M. S. Farrell, J. Frank, L. Franke, R. Freedman, N. B. Freimer, M. Friedl, J. I. Friedman, M. Fromer, G. Genovese, L. Georgieva, I. Giegling, P. Giusti-Rodriguez, S. Godard, J. I. Goldstein, V. Golimbet, S. Gopal, J. Gratten, L. de Haan, C. Hammer, M. L. Hamshere, M. Hansen, T. Hansen, V. Haroutunian, A. M. Hartmann, F. A. Henskens, S. Herms, J. N. Hirschhorn, P. Hoffmann, A. Hofman, M. V. Hollegaard, D. M. Hougaard, M. Ikeda, I. Joa, A. Julia, R. S. Kahn, L. Kalaydjieva, S. Karachanak-Yankova, J. Karjalainen, D. Kavanagh, M. C. Keller, J. L. Kennedy, A. Khrunin, Y. Kim, J. Klovins, J. A. Knowles, B. Konte, V. Kucinskas, Z. Ausrele Kucinskiene, H. Kuzelova-Ptackova, A. K. Kahler, C. Laurent, J. L. Keong, S. H. Lee, S. E. Legge, B. Lerer, M. Li, T. Li, K. Y. Liang, J. Lieberman, S. Limborska, C. M. Loughland, J. Lubinski, J. Lonnqvist, M. Macek, P. K. Magnusson, B. S. Maher, W. Maier, J. Mallet, S. Marsal, M. Mattheisen, M. Mattingsdal, R. W. McCarley, C. McDonald, A. M. McIntosh, S. Meier, C. J. Meijer, B. Melegh, I. Melle, R. I. Mesholam-Gately, A. Metspalu, P. T. Michie, L. Milani, V. Milanova, Y. Mokrab, D. W. Morris, O. Mors, K. C. Murphy, R. M. Murray, I. Myin-Germeys, B. Muller-Myhsok, M. Nelis, I. Nenadic, D. A. Nertney, G. Nestadt, K. K. Nicodemus, L. Nikitina-Zake, L. Nisenbaum, A. Nordin, E. O'Callaghan, C. O'Dushlaine, F. A. O'Neill, S. Y. Oh, A. Olincy, L. Olsen, J. Van Os, C. Pantelis, G. N. Papadimitriou, S. Papiol, E. Parkhomenko, M. T. Pato, T. Paunio, M. Pejovic-Milovancevic, D. O. Perkins, O. Pietilainen, J. Pimm, A. J. Pocklington, J. Powell, A. Price, A. E. Pulver, S. M. Purcell, D. Quested, H. B. Rasmussen, A. Reichenberg, M. A. Reimers, A. L. Richards, J. L. Roffman, P. Roussos, D. M. Ruderfer, V. Salomaa, A. R. Sanders, U. Schall, C. R. Schubert, T. G. Schulze, S. G. Schwab, E. M. Scolnick, R. J. Scott, L. J. Seidman, J. Shi, E. Sigurdsson, T. Silagadze, J. M. Silverman, K. Sim, P. Slominsky, J. W. Smoller, H. C. So, C. A. Spencer, E. A. Stahl, H. Stefansson, S. Steinberg, E. Stogmann, R. E. Straub, E. Strengman, J. Strohmaier, T. S. Stroup, M. Subramaniam, J. Suvisaari, D. M. Svrakic, J. P. Szatkiewicz, E. Soderman, S. Thirumalai, D. Toncheva, S. Tosato, J. Veijola, J. Waddington, D. Walsh, D. Wang, Q. Wang, B. T. Webb, M. Weiser, D. B. Wildenauer, N. M. Williams, S. Williams, S. H. Witt, A. R. Wolen, E. H. Wong, B. K. Wormley, H. S. Xi, C. C. Zai, X. Zheng, F. Zimprich, N. R. Wray, K. Stefansson, P. M. Visscher, R. Adolfsson, O. A. Andreassen, D. H. Blackwood, E. Bramon, J. D. Buxbaum, A. D. B?rglum, S. Cichon, A. Darvasi, E. Domenici, H. Ehrenreich, T. Esko, P. V. Gejman, M. Gill, H. Gurling, C. M. Hultman, N. Iwata, A. V. Jablensky, E. G. Jonsson, K. S. Kendler, G. Kirov, J. Knight, T. Lencz, D. F. Levinson, Q. S. Li, J. Liu, A. K. Malhotra, S. A. McCarroll, A. McQuillin, J. L. Moran, P. B. Mortensen, B. J. Mowry, M. M. Nothen, R. A. Ophoff, M. J. Owen, A. Palotie, C. N. Pato, T. L. Petryshen, D. Posthuma, M. Rietschel, B. P. Riley, D. Rujescu, P. C. Sham, P. Sklar, D. St Clair, D. R. Weinberger, J. R. Wendland, T. Werge, M. J. Daly, P. F. Sullivan, and M. C. O'Donovan. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510):421–427, Jul 2014. PMCID: PMC4112379.

[5] D. Altshuler, M. J. Daly, and E. S. Lander. Genetic mapping in human disease. *Science*, 322(5903):881–888, Nov 2008. PMCID: PMC2694957.

[6] J. Hardy and A. Singleton. Genomewide association studies and human disease. *N. Engl. J. Med.*, 360(17):1759–1768, Apr 2009. PMCID: PMC3422859.

[7] B. Devlin and S. W. Scherer. Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Dev.*, 22(3):229–237, Jun 2012.

[8] J. A. Veltman and H. G. Brunner. De novo mutations in human genetic disease. *Nat. Rev. Genet.*, 13(8):565–575, Aug 2012.

[9] C. S. Ku, C. Polychronakos, E. K. Tan, N. Naidoo, Y. Pawitan, D. H. Roukos, M. Mort, and D. N. Cooper. A new paradigm emerges from the study of de novo mutations in the context of neurodevelopmental disease. *Mol. Psychiatry*, 18(2):141–153, Feb 2013.

[10] S. De Rubeis, X. He, and et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515(7526):209–215, Nov 2014. PMCID: PMC4402723.

[11] I. Iossifov, B. J. O'Roak, S. J. Sanders, M. Ronemus, N. Krumm, D. Levy, H. A. Stessman, K. T. Witherspoon, L. Vives, K. E. Patterson, J. D. Smith, B. Paeper, D. A. Nickerson, J. Dea, S. Dong, L. E. Gonzalez, J. D. Mandell, S. M. Mane, M. T. Murtha, C. A. Sullivan, M. F. Walker, Z. Waqar, L. Wei, A. J. Willsey, B. Yamrom, Y. H. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M. C. Schatz, K. Ye, W. R. McCombie, J. Shendure, E. E. Eichler, M. W. State, and M. Wigler. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526):216–221, Nov 2014. PMCID: PMC4313871.

[12] C. Gilissen, J. Y. Hehir-Kwa, D. T. Thung, M. van de Vorst, B. W. van Bon, M. H. Willemsen, M. Kwint, I. M. Janssen, A. Hoischen, A. Schenck, R. Leach, R. Klein, R. Tearle, T. Bo, R. Pfundt, H. G. Yntema, B. B. de Vries, T. Kleefstra, H. G. Brunner, L. E. Vissers, and J. A. Veltman. Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 511(7509):344–347, Jul 2014.

[13] M. Fromer, A. J. Pocklington, D. H. Kavanagh, H. J. Williams, S. Dwyer, P. Gormley, L. Georgieva, E. Rees, P. Palta, D. M. Ruderfer, N. Carrera, I. Humphreys, J. S. Johnson, P. Roussos, D. D. Barker, E. Banks, V. Milanova, S. G. Grant, E. Hannon, S. A. Rose, K. Chambert, M. Mahajan, E. M. Scolnick, J. L. Moran, G. Kirov, A. Palotie, S. A. McCarroll, P. Holmans, P. Sklar, M. J. Owen, S. M. Purcell, and M. C. O'Donovan. De novo mutations in schizophrenia implicate synaptic networks. *Nature*, 506(7487):179–184, Feb 2014. PMCID: PMC4237002.

[14] A. S. Allen, S. F. Berkovic, P. Cossette, N. Delanty, D. Dlugos, E. E. Eichler, M. P. Epstein, T. Glauser, D. B. Goldstein, Y. Han, E. L. Heinzen, Y. Hitomi, K. B. Howell, M. R. Johnson, R. Kuzniecky, D. H. Lowenstein, Y. F. Lu, M. R. Madou, A. G. Marson, H. C. Mefford, S. Esmaeeli Nieh, T. J. O'Brien, R. Ottman, S. Petrovski, A. Poduri, E. K. Ruzzo, I. E. Scheffer, E. H. Sherr, C. J. Yuskaitis, B. Abou-Khalil, B. K. Alldredge, J. F. Bautista, S. F. Berkovic, A. Boro, G. D. Cascino, D. Consalvo, P. Crumrine, O. Devinsky, D. Dlugos, M. P. Epstein, M. Fiol, N. B. Fountain, J. French, D. Friedman, E. B. Geller, T. Glauser, S. Glynn, S. R. Haut, J. Hayward, S. L. Helmers, S. Joshi, A. Kanner, H. E. Kirsch, R. C. Knowlton, E. H. Kossoff, R. Kuperman, R. Kuzniecky, D. H. Lowenstein, S. M. McGuire, P. V. Motika, E. J. Novotny, R. Ottman, J. M. Paolicchi, J. M. Parent, K. Park, A. Poduri, I. E. Scheffer, R. A. Shellhaas, E. H. Sherr, J. J. Shih, R. Singh, J. Sirven, M. C. Smith, J. Sullivan, L. Lin Thio, A. Venkat, E. P. Vining, G. K. Von Allmen, J. L. Weisenberg, P. Widdess-Walsh, and M. R. Winawer. De novo mutations in epileptic encephalopathies. *Nature*, 501(7466):217–221, Sep 2013. PMCID: PMC3773011.

[15] S. Zaidi, M. Choi, H. Wakimoto, L. Ma, J. Jiang, J. D. Overton, A. Romano-Adesman, R. D. Bjornson, R. E. Breitbart, K. K. Brown, N. J. Carriero, Y. H. Cheung, J. Deanfield, S. DePalma, K. A. Fakhro, J. Glessner, H. Hakonarson, M. J. Italia, J. R. Kaltman, J. Kaski, R. Kim, J. K. Kline, T. Lee, J. Leipzig, A. Lopez, S. M. Mane, L. E. Mitchell, J. W. Newburger, M. Parfenov, I. Pe'er, G. Porter, A. E. Roberts, R. Sachidanandam, S. J. Sanders, H. S. Seiden, M. W. State, S. Subramanian, I. R. Tikhonova, W. Wang, D. Warburton, P. S. White, I. A. Williams, H. Zhao, J. G. Seidman, M. Brueckner, W. K. Chung, B. D. Gelb, E. Goldmuntz, C. E. Seidman, and R. P. Lifton. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*, 498(7453):220–223, Jun 2013. PMCID: PMC3706629.

[16] S. J. Sanders, X. He, A. J. Willsey, A. G. Ercan-Sencicek, K. E. Samocha, A. E. Cicek, M. T. Murtha, V. H. Bal, S. L. Bishop, S. Dong, A. P. Goldberg, C. Jinlu, J. F. Keaney, L. Klei, J. D. Mandell, D. Moreno-De-Luca, C. S. Poultney, E. B. Robinson, L. Smith, T. Solli-Nowlan, M. Y. Su, N. A. Teran, M. F. Walker, D. M. Werling, A. L. Beaudet, R. M. Cantor, E. Fombonne, D. H. Geschwind, D. E. Grice, C. Lord, J. K. Lowe, S. M. Mane, D. M. Martin, E. M. Morrow, M. E. Talkowski, J. S. Sutcliffe, C. A. Walsh, T. W. Yu, D. H. Ledbetter, C. L. Martin, E. H. Cook, J. D. Buxbaum, M. J. Daly, B. Devlin, K. Roeder, and M. W. State. Insights into Autism

Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, 87(6):1215–1233, Sep 2015.

[17] S. J. Sanders, M. T. Murtha, A. R. Gupta, J. D. Murdoch, M. J. Raubeson, A. J. Willsey, A. G. Ercan-Sencicek, N. M. DiLullo, N. N. Parikshak, J. L. Stein, M. F. Walker, G. T. Ober, N. A. Teran, Y. Song, P. El-Fishawy, R. C. Murtha, M. Choi, J. D. Overton, R. D. Bjornson, N. J. Carriero, K. A. Meyer, K. Bilguvar, S. M. Mane, N. Sestan, R. P. Lifton, M. Gunel, K. Roeder, D. H. Geschwind, B. Devlin, and M. W. State. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485(7397):237–241, May 2012. PMCID: PMC3667984.

[18] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A. Stamatoyannopoulos. Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337(6099):1190–1195, Sep 2012. PMCID: PMC3771521.

[19] Y. H. Jiang, R. K. Yuen, X. Jin, M. Wang, N. Chen, X. Wu, J. Ju, J. Mei, Y. Shi, M. He, G. Wang, J. Liang, Z. Wang, D. Cao, M. T. Carter, C. Chrysler, I. E. Drmic, J. L. Howe, L. Lau, C. R. Marshall, D. Merico, T. Nalpathamkalam, B. Thiruvahindrapuram, A. Thompson, M. Uddin, S. Walker, J. Luo, E. Anagnostou, L. Zwaigenbaum, R. H. Ring, J. Wang, C. Lajonchere, J. Wang, A. Shih, P. Szatmari, H. Yang, G. Dawson, Y. Li, and S. W. Scherer. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.*, 93(2):249–263, Aug 2013. PMCID: PMC3738824.

[20] R. K. Yuen, B. Thiruvahindrapuram, D. Merico, S. Walker, K. Tammimies, N. Hoang, C. Chrysler, T. Nalpathamkalam, G. Pellecchia, Y. Liu, M. J. Gazzellone, L. D'Abate, E. Deneault, J. L. Howe, R. S. Liu, A. Thompson, M. Zarrei, M. Uddin, C. R. Marshall, R. H. Ring, L. Zwaigenbaum, P. N. Ray, R. Weksberg, M. T. Carter, B. A. Fernandez, W. Roberts, P. Szatmari, and S. W. Scherer. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.*, 21(2):185–191, Feb 2015.

[21] S. Girirajan, C. D. Campbell, and E. E. Eichler. Human copy number variation and complex genetic disease. *Annu. Rev. Genet.*, 45:203–226, 2011.

[22] D. Pinto, A. T. Pagnamenta, L. Klei, R. Anney, D. Merico, R. Regan, J. Conroy, T. R. Magalhaes, C. Correia, B. S. Abrahams, J. Almeida, E. Bacchelli, G. D. Bader, A. J. Bailey, G. Baird, A. Battaglia, T. Berney, N. Bolshakova, S. Bolte, P. F. Bolton, T. Bourgeron, S. Brennan, J. Brian, S. E. Bryson, A. R. Carson, G. Casallo, J. Casey, B. H. Chung, L. Cochrane, C. Corsello, E. L. Crawford, A. Crossett, C. Cytrynbaum, G. Dawson, M. de Jonge, R. Delorme, I. Drmic, E. Duketis, F. Duque, A. Estes, P. Farrar, B. A. Fernandez, S. E. Folstein, E. Fombonne, C. M. Freitag, J. Gilbert, C. Gillberg, J. T. Glessner, J. Goldberg, A. Green, J. Green, S. J. Guter, H. Hakonarson, E. A. Heron, M. Hill, R. Holt, J. L. Howe, G. Hughes, V. Hus, R. Igliozzi, C. Kim, S. M. Klauck, A. Kolevzon, O. Korvatska, V. Kustanovich, C. M. Lajonchere, J. A. Lamb, M. Laskawiec, M. Leboyer, A. Le Couteur, B. L. Leventhal, A. C. Lionel, X. Q. Liu, C. Lord, L. Lotspeich, S. C. Lund, E. Maestrini, W. Mahoney, C. Mantoulan, C. R. Marshall, H. McConachie, C. J. McDougle, J. McGrath, W. M. McMahon, A. Merikangas, O. Migita, N. J. Minshew, G. K. Mirza, J. Munson, S. F. Nelson, C. Noakes, A. Noor, G. Nygren, G. Oliveira, K. Papanikolaou, J. R. Parr, B. Parrini, T. Paton, A. Pickles, M. Pilorge, J. Piven, C. P. Ponting, D. J. Posey, A. Poustka, F. Poustka, A. Prasad, J. Ragoussis, K. Renshaw, J. Rickaby, W. Roberts, K. Roeder, B. Roge, M. L. Rutter, L. J. Bierut, J. P. Rice, J. Salt, K. Sansom, D. Sato, R. Segurado, A. F. Sequeira, L. Senman, N. Shah, V. C. Sheffield, L. Soorya, I. Sousa, O. Stein, N. Sykes, V. Stoppioni, C. Strawbridge, R. Tancredi, K. Tansey, B. Thiruvahindrapduram, A. P. Thompson, S. Thomson, A. Tryfon, J. Tsiantis, H. Van Engeland, J. B. Vincent, F. Volkmar, S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Wittemeyer, S. Wood, J. Wu, B. L. Yaspan, D. Zurawiecki, L. Zwaigenbaum, J. D. Buxbaum, R. M. Cantor, E. H. Cook, H. Coon, M. L. Cuccaro, B. Devlin, S. Ennis, L. Gallagher, D. H. Geschwind, M. Gill, J. L. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger, A. D. Paterson, M. A. Pericak-Vance, G. D. Schellenberg, P. Szatmari, A. M. Vicente, V. J. Vieland, E. M. Wijsman, S. W. Scherer, J. S. Sutcliffe, and C. Betancur. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, 466(7304):368–372, Jul 2010.

[23] S. J. Sanders, A. G. Ercan-Sencicek, V. Hus, R. Luo, M. T. Murtha, D. Moreno-De-Luca, S. H. Chu, M. P. Moreau, A. R. Gupta, S. A. Thomson, C. E. Mason, K. Bilguvar, P. B. Celestino-Soper, M. Choi, E. L. Crawford, L. Davis, N. R. Wright, R. M. Dhodapkar, M. DiCola, N. M. DiLullo, T. V. Fernandez, V. Fielding-Singh, D. O. Fishman, S. Frahm, R. Garagaloyan, G. S. Goh, S. Kammela, L. Klei, J. K. Lowe, S. C. Lund, A. D. McGrew, K. A. Meyer, W. J. Moffat, J. D. Murdoch, B. J. O'Roak, G. T. Ober, R. S. Pottenger, M. J. Raubeson, Y. Song, Q. Wang, B. L. Yaspan, T. W. Yu, I. R. Yurkiewicz, A. L. Beaudet, R. M. Cantor, M. Curland, D. E. Grice, M. Gunel, R. P. Lifton, S. M. Mane, D. M. Martin, C. A. Shaw, M. Sheldon, J. A. Tischfield, C. A. Walsh, E. M. Morrow, D. H. Ledbetter, E. Fombonne, C. Lord, C. L. Martin, A. I. Brooks, J. S. Sutcliffe, E. H. Cook, D. Geschwind, K. Roeder, B. Devlin, and M. W. State. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, 70(5):863–885, Jun 2011.

[24] C. Alkan, B. P. Coe, and E. E. Eichler. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.*, 12(5):363–376, May 2011.

[25] S. Raychaudhuri, J. M. Korn, S. A. McCarroll, D. Altshuler, P. Sklar, S. Purcell, M. J. Daly, S. Purcell, J. Stone, S. Bergen, C. O'Dushlaine, D. Ruderfer, P. Sklar, E. Scolnick, K. Chambert, M. O'Donovan, G. Kirov, N. Craddock, P. Holmans, N. Williams, L. Georgieva, I. Nikolov, N. Norton, H. Williams, D. Toncheva, V. Milanova, M. Owen, C. Hultman, P. Lichtenstein, E. Thelander, P. Sullivan, D. Morris, E. Kenny, J. Waddington, M. Gill, A. Corvin, A. McQuillin, K. Choudhury, S. Datta, J. Pimm, S. Thirumalai, V. Puri, R. Krasucki, J. Lawrence, D. Quested, N. Bass, D. Curtis, H. Gurling, C. Crombie, G. Fraser, N. Kwan, N. Walker, D. St Clair, D. Blackwood, W. Muir, K. McGhee, A. Maclean, M. Van Beck, P. Visscher, S. Macgregor, N. Wray, M. T. Pato, H. Medeiros, F. Middleton, C. Carvalho, C. Morley, A. Fanous, D. Conti, J. Knowles, C. P. Ferreira, A. Macedo, M. H. Azevedo, and C. N. Pato. Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet.*, 6(9):e1001097, Sep 2010.

[26] D. Pinto, E. Delaby, D. Merico, M. Barbosa, A. Merikangas, L. Klei, B. Thiruvahindrapuram, X. Xu, R. Ziman, Z. Wang, J. A. Vorstman, A. Thompson, R. Regan, M. Pilorge, G. Pellecchia, A. T. Pagnamenta, B. Oliveira, C. R. Marshall, T. R. Magalhaes, J. K. Lowe, J. L. Howe, A. J. Griswold, J. Gilbert, E. Duketis, B. A. Dombroski, M. V. De Jonge, M. Cuccaro, E. L. Crawford, C. T. Correia, J. Conroy, I. C. Conceicao, A. G. Chiocchetti, J. P. Casey, G. Cai, C. Cabrol, N. Bolshakova, E. Bacchelli, R. Anney, S. Gallinger, M. Cotterchio, G. Casey, L. Zwaigenbaum, K. Wittemeyer, K. Wing, S. Wallace, H. van Engeland, A. Tryfon, S. Thomson, L. Soorya, B. Roge, W. Roberts, F. Poustka, S. Mouga, N. Minshew, L. A. McInnes, S. G. McGrew, C. Lord, M. Leboyer, A. S. Le Couteur, A. Kolevzon, P. Jimenez Gonzalez, S. Jacob, R. Holt, S. Guter, J. Green, A. Green, C. Gillberg, B. A. Fernandez, F. Duque, R. Delorme, G. Dawson, P. Chaste, C. Cafe, S. Brennan, T. Bourgeron, P. F. Bolton, S. Bolte, R. Bernier, G. Baird, A. J. Bailey, E. Anagnostou, J. Almeida, E. M. Wijsman, V. J. Vieland, A. M. Vicente, G. D. Schellenberg, M. Pericak-Vance, A. D. Paterson, J. R. Parr, G. Oliveira, J. I. Nurnberger, A. P. Monaco, E. Maestrini, S. M. Klauck, H. Hakonarson, J. L. Haines, D. H. Geschwind, C. M. Freitag, S. E. Folstein, S. Ennis, H. Coon, A. Battaglia, P. Szatmari, J. S. Sutcliffe, J. Hallmayer, M. Gill, E. H. Cook, J. D. Buxbaum, B. Devlin, L. Gallagher, C. Betancur, and S. W. Scherer. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.*, 94(5):677–694, May 2014. PMCID: PMC4067558.

[27] A. J. Pocklington, E. Rees, J. T. Walters, J. Han, D. H. Kavanagh, K. D. Chambert, P. Holmans, J. L. Moran, S. A. McCarroll, G. Kirov, M. C. O'Donovan, and M. J. Owen. Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. *Neuron*, 86(5):1203–1214, Jun 2015. PMCID: PMC4460187.

[28] D. L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M. E. Dolan, and N. J. Cox. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.*, 6(4):e1000888, Apr 2010.

[29] K. G. Ardlie, D. S. Deluca, A. V. Segre, T. J. Sullivan, T. R. Young, E. T. Gelfand, C. A. Trowbridge, J. B. Maller, T. Tukiainen, M. Lek, L. D. Ward, P. Kheradpour, B. Iriarte, Y. Meng, C. D. Palmer, T. Esko, W. Winckler, J. N. Hirschhorn, M. Kellis, D. G. MacArthur, G. Getz, A. A. Shabalin, G. Li, Y. H. Zhou, A. B. Nobel,

I. Rusyn, F. A. Wright, T. Lappalainen, P. G. Ferreira, H. Ongen, M. A. Rivas, A. Battle, S. Mostafavi, J. Monlong, M. Sammeth, M. Mele, F. Reverter, J. M. Goldmann, D. Koller, R. Guigo, M. I. McCarthy, E. T. Dermitzakis, E. R. Gamazon, H. K. Im, A. Konkashbaev, D. L. Nicolae, N. J. Cox, T. Flutre, X. Wen, M. Stephens, J. K. Pritchard, Z. Tu, B. Zhang, T. Huang, Q. Long, L. Lin, J. Yang, J. Zhu, J. Liu, A. Brown, B. Mestichelli, D. Tidwell, E. Lo, M. Salvatore, S. Shad, J. A. Thomas, J. T. Lonsdale, M. T. Moser, B. M. Gillard, E. Karasik, K. Ramsey, C. Choi, B. A. Foster, J. Syron, J. Fleming, H. Magazine, R. Hasz, G. D. Walters, J. P. Bridge, M. Miklos, S. Sullivan, L. K. Barker, H. M. Traino, M. Mosavel, L. A. Siminoff, D. R. Valley, D. C. Rohrer, S. D. Jewell, P. A. Branton, L. H. Sobin, M. Barcus, L. Qi, J. McLean, P. Hariharan, K. S. Um, S. Wu, D. Tabor, C. Shive, A. M. Smith, S. A. Buia, A. H. Undale, K. L. Robinson, N. Roche, K. M. Valentino, A. Britton, R. Burges, D. Bradbury, K. W. Hambright, J. Seleski, G. E. Korzeniewski, K. Erickson, Y. Marcus, J. Tejada, M. Taherian, C. Lu, M. Basile, D. C. Mash, S. Volpi, J. P. Struewing, G. F. Temple, J. Boyer, D. Colantuoni, R. Little, S. Koester, L. J. Carithers, H. M. Moore, P. Guan, C. Compton, S. J. Sawyer, J. P. Demchok, J. B. Vaught, C. A. Rabiner, N. C. Lockhart, K. G. Ardlie, G. Getz, F. A. Wright, M. Kellis, S. Volpi, and E. T. Dermitzakis. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235):648–660, May 2015.

[30] W. Cookson, L. Liang, G. Abecasis, M. Moffatt, and M. Lathrop. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.*, 10(3):184–194, Mar 2009.

[31] S. B. Montgomery and E. T. Dermitzakis. From expression QTLs to personalized transcriptomics. *Nat. Rev. Genet.*, 12(4):277–282, Apr 2011.

[32] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.*, 47(9):1091–1098, Sep 2015.

[33] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma, F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. Lehtimaki, E. Raitoharju, M. Kahonen, I. Seppala, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.*, 48(3):245–252, Mar 2016.

[34] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.*, 48(5):481–487, May 2016.

[35] A. L. Price, A. Helgason, G. Thorleifsson, S. A. McCarroll, A. Kong, and K. Stefansson. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet.*, 7(2):e1001317, Feb 2011.

[36] E. Grundberg, K. S. Small, A. K. Hedman, A. C. Nica, A. Buil, S. Keildson, J. T. Bell, T. P. Yang, E. Meduri, A. Barrett, J. Nisbett, M. Sekowska, A. Wilk, S. Y. Shin, D. Glass, M. Travers, J. L. Min, S. Ring, K. Ho, G. Thorleifsson, A. Kong, U. Thorsteindottir, C. Ainali, A. S. Dimas, N. Hassanali, C. Ingle, D. Knowles, M. Krestyaninova, C. E. Lowe, P. Di Meglio, S. B. Montgomery, L. Parts, S. Potter, G. Surdulescu, L. Tsaprouni, S. Tsoka, V. Bataille, R. Durbin, F. O. Nestle, S. O'Rahilly, N. Soranzo, C. M. Lindgren, K. T. Zondervan, K. R. Ahmadi, E. E. Schadt, K. Stefansson, G. D. Smith, M. I. McCarthy, P. Deloukas, E. T. Dermitzakis, and T. D. Spector. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.*, 44(10):1084–1089, Oct 2012.

[37] A. Visel, E. M. Rubin, and L. A. Pennacchio. Genomic views of distant-acting enhancers. *Nature*, 461(7261):199–205, Sep 2009. PMCID: PMC2923221.

[38] A. S. Nord, K. Pattabiraman, A. Visel, and J. L. Rubenstein. Genomic perspectives of transcriptional regulation in forebrain development. *Neuron*, 85(1):27–47, Jan 2015. PMCID: PMC4438709.

[39] X. He, C. K. Fuller, Y. Song, Q. Meng, B. Zhang, X. Yang, and H. Li. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am. J. Hum. Genet.*, 92(5):667–680, May 2013. PMCID: PMC3644637.

[40] H. J. Westra, M. J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, M. W. Christiansen, B. P. Fairfax, K. Schramm, J. E. Powell, A. Zhernakova, D. V. Zhernakova, J. H. Veldink, L. H. Van den Berg, J. Karjalainen, S. Withoff, A. G. Uitterlinden, A. Hofman, F. Rivadeneira, P. A. 't Hoen, E. Reinmaa, K. Fischer, M. Nelis, L. Milani, D. Melzer, L. Ferrucci, A. B. Singleton, D. G. Hernandez, M. A. Nalls, G. Homuth, M. Nauck, D. Radke, U. Volker, M. Perola, V. Salomaa, J. Brody, A. Suchy-Dicey, S. A. Gharib, D. A. Enquobahrie, T. Lumley, G. W. Montgomery, S. Makino, H. Prokisch, C. Herder, M. Roden, H. Grallert, T. Meitinger, K. Strauch, Y. Li, R. C. Jansen, P. M. Visscher, J. C. Knight, B. M. Psaty, S. Ripatti, A. Teumer, T. M. Frayling, A. Metspalu, J. B. van Meurs, and L. Franke. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, 45(10):1238–1243, Oct 2013.

[41] A. A. Vinkhuyzen, N. R. Wray, J. Yang, M. E. Goddard, and P. M. Visscher. Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annu. Rev. Genet.*, 47:75–95, 2013.

[42] A. A. Pai, J. K. Pritchard, and Y. Gilad. The genetic and mechanistic basis for variation in gene regulation. *PLoS Genet.*, 11(1):e1004857, Jan 2015.

[43] Autism Spectrum Disorder: Data and Statistics. `http://www.cdc.gov/ncbddd/autism/data.html`.

[44] A. L. Richards, L. Jones, V. Moskvina, G. Kirov, P. V. Gejman, D. F. Levinson, A. R. Sanders, S. Purcell, P. M. Visscher, N. Craddock, M. J. Owen, P. Holmans, M. C. O'Donovan, P. V. Gejman, A. R. Sanders, J. Duan, D. F. Levinson, N. G. Buccola, B. J. Mowry, R. Freedman, F. Amin, D. W. Black, J. M. Silverman, W. J. Byerley, C. R. Cloninger, M. C. O'Donovan, G. K. Kirov, N. J. Craddock, P. A. Holmans, N. M. Williams, L. Georgieva, I. Nikolov, N. Norton, H. Williams, D. Toncheva, V. Milanova, M. J. Owen, C. M. Hultman, P. Lichtenstein, E. F. Thelander, P. Sullivan, D. W. Morris, C. T. O'Dushlaine, E. Kenny, E. M. Quinn, M. Gill, A. Corvin, A. McQuillin, K. Choudhury, S. Datta, J. Pimm, S. Thirumalai, V. Puri, R. Krasucki, J. Lawrence, D. Quested, N. Bass, H. Gurling, C. Crombie, G. Fraser, S. L. Kuan, N. Walker, D. S. Clair, D. H. Blackwood, W. J. Muir, K. A. McGhee, B. Pickard, P. Malloy, A. W. Maclean, M. V. Beck, N. R. Wray, S. Macgregor, P. M. Visscher, M. T. Pato, H. Medeiros, F. Middleton, C. Carvalho, C. Morley, A. Fanous, D. Conti, J. A. Knowles, C. P. Ferreira, A. Macedo, M. H. Azevedo, C. N. Pato, J. L. Stone, A. N. Kirby, M. A. Ferreira, M. J. Daly, S. M. Purcell, J. L. Stone, K. Chambert, D. M. Ruderfer, F. Kuruvilla, S. B. Gabriel, K. Ardlie, J. L. Moran, E. M. Scolnick, and P. Sklar. Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol. Psychiatry*, 17(2):193–201, Feb 2012.

[45] P. Roussos, A. C. Mitchell, G. Voloudakis, J. F. Fullard, V. M. Pothula, J. Tsang, E. A. Stahl, A. Georgakopoulos, D. M. Ruderfer, A. Charney, Y. Okada, K. A. Siminovitch, J. Worthington, L. Padyukov, L. Klareskog, P. K. Gregersen, R. M. Plenge, S. Raychaudhuri, M. Fromer, S. M. Purcell, K. J. Brennand, N. K. Robakis, E. E. Schadt, S. Akbarian, and P. Sklar. A role for noncoding variation in schizophrenia. *Cell Rep*, 9(4):1417–1429, Nov 2014.

[46] X. He, S. J. Sanders, L. Liu, S. De Rubeis, E. T. Lim, J. S. Sutcliffe, G. D. Schellenberg, R. A. Gibbs, M. J. Daly, J. D. Buxbaum, M. W. State, B. Devlin, and K. Roeder. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet.*, 9(8):e1003671, 2013. PMCID: PMC3744441.

[47] L. Liu, J. Lei, S. J. Sanders, A. J. Willsey, Y. Kou, A. E. Cicek, L. Klei, C. Lu, X. He, M. Li, R. A. Muhle, A. Ma'ayan, J. P. Noonan, N. Sestan, K. A. McFadden, M. W. State, J. D. Buxbaum, B. Devlin, and K. Roeder. DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. *Mol Autism*, 5(1):22, 2014. PMCID: PMC4016412.

[48] X. He, M. A. Samee, C. Blatti, and S. Sinha. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Comput. Biol.*, 6(9), 2010. PMCID: PMC2940721.

[49] S. Zhong, X. He, and Z. Bar-Joseph. Predicting tissue specific transcription factor binding sites. *BMC Genomics*, 14:796, 2013. PMCID: PMC3898213.

[50] X. He, X. Ling, and S. Sinha. Alignment and prediction of cis-regulatory modules based on a probabilistic model of evolution. *PLoS Comput. Biol.*, 5(3):e1000299, Mar 2009. PMCID: PMC2657044.

[51] X. He, T. S. Duque, and S. Sinha. Evolutionary origins of transcription factor binding site clusters. *Mol. Biol. Evol.*, 29(3):1059–1070, Mar 2012. PMCID: PMC3278477.

[52] X. He, A. E. Cicek, Y. Wang, M. H. Schulz, H. S. Le, and Z. Bar-Joseph. De novo ChIP-seq analysis. *Genome Biol.*, 16(1):205, 2015.

[53] X. He and M. H. Goldwasser. Identifying conserved gene clusters in the presence of homology families. *J. Comput. Biol.*, 12(6):638–656, 2005.

[54] J. K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, Jun 2000. PMCID: PMC1461096.

[55] M. Stephens, N. J. Smith, and P. Donnelly. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, 68(4):978–989, Apr 2001. PMCID: PMC1275651.

[56] B. M. Neale, Y. Kou, L. Liu, A. Ma'ayan, K. E. Samocha, A. Sabo, C. F. Lin, C. Stevens, L. S. Wang, V. Makarov, P. Polak, S. Yoon, J. Maguire, E. L. Crawford, N. G. Campbell, E. T. Geller, O. Valladares, C. Schafer, H. Liu, T. Zhao, G. Cai, J. Lihm, R. Dannenfelser, O. Jabado, Z. Peralta, U. Nagaswamy, D. Muzny, J. G. Reid, I. Newsham, Y. Wu, L. Lewis, Y. Han, B. F. Voight, E. Lim, E. Rossin, A. Kirby, J. Flannick, M. Fromer, K. Shakir, T. Fennell, K. Garimella, E. Banks, R. Poplin, S. Gabriel, M. DePristo, J. R. Wimbish, B. E. Boone, S. E. Levy, C. Betancur, S. Sunyaev, E. Boerwinkle, J. D. Buxbaum, E. H. Cook, B. Devlin, R. A. Gibbs, K. Roeder, G. D. Schellenberg, J. S. Sutcliffe, and M. J. Daly. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, 485(7397):242–245, May 2012. PMCID: PMC3613847.

[57] B. J. O'Roak, L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D. Smith, E. H. Turner, I. B. Stanaway, B. Vernot, M. Malig, C. Baker, B. Reilly, J. M. Akey, E. Borenstein, M. J. Rieder, D. A. Nickerson, R. Bernier, J. Shendure, and E. E. Eichler. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, 485(7397):246–250, May 2012. PMCID: PMC3350576.

[58] K. E. Samocha, E. B. Robinson, S. J. Sanders, C. Stevens, A. Sabo, L. M. McGrath, J. A. Kosmicki, K. Rehnstrom, S. Mallick, A. Kirby, D. P. Wall, D. G. MacArthur, S. B. Gabriel, M. DePristo, S. M. Purcell, A. Palotie, E. Boerwinkle, J. D. Buxbaum, E. H. Cook, R. A. Gibbs, G. D. Schellenberg, J. S. Sutcliffe, B. Devlin, K. Roeder, B. M. Neale, and M. J. Daly. A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.*, 46(9):944–950, Sep 2014. PMCID: PMC4222185.

[59] B. Li and S. M. Leal. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.*, 83(3):311–321, Sep 2008.

[60] S. Lee, G. R. Abecasis, M. Boehnke, and X. Lin. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.*, 95(1):5–23, Jul 2014.

[61] M. C. Wu, S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.*, 89(1):82–93, Jul 2011.

[62] A. Kong, M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W. S. Wong, G. Sigurdsson, G. B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D. F. Gudbjartsson, A. Helgason, O. T. Magnusson, U. Thorsteinsdottir, and K. Stefansson. Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, 488(7412):471–475, Aug 2012. PMCID: PMC3548427.

[63] J. J. Michaelson, Y. Shi, M. Gujral, H. Zheng, D. Malhotra, X. Jin, M. Jian, G. Liu, D. Greer, A. Bhandari, W. Wu, R. Corominas, A. Peoples, A. Koren, A. Gore, S. Kang, G. N. Lin, J. Estabillo, T. Gadomski, B. Singh,

K. Zhang, N. Akshoomoff, C. Corsello, S. McCarroll, L. M. Iakoucheva, Y. Li, J. Wang, and J. Sebat. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell*, 151(7):1431–1442, Dec 2012. PMCID: PMC3712641.

[64] W. S. Wong, B. D. Solomon, D. L. Bodian, P. Kothiyal, G. Eley, K. C. Huddleston, R. Baker, D. C. Thach, R. K. Iyer, J. G. Vockley, and J. E. Niederhuber. New observations on maternal age effect on germline de novo mutations. *Nat Commun*, 7:10486, 2016.

[65] S. K. Reilly, J. Yin, A. E. Ayoub, D. Emera, J. Leng, J. Cotney, R. Sarro, P. Rakic, and J. P. Noonan. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science*, 347(6226):1155–1159, Mar 2015. PMCID: PMC4426903.

[66] A. Mathelier, O. Fornes, D. J. Arenillas, C. Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, and W. W. Wasserman. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, 44(D1):D110–115, Jan 2016.

[67] D. J. Gaffney, J. B. Veyrieras, J. F. Degner, R. Pique-Regi, A. A. Pai, G. E. Crawford, M. Stephens, Y. Gilad, and J. K. Pritchard. Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.*, 13(1):R7, 2012. PMCID: PMC3334587.

[68] P. Carbonetto and M. Stephens. Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease. *PLoS Genet.*, 9(10):e1003770, 2013. PMCID: PMC3789883.

[69] T. Flutre, X. Wen, J. Pritchard, and M. Stephens. A statistical framework for joint eQTL analysis in multiple tissues. *PLoS Genet.*, 9(5):e1003486, May 2013. PMCID: PMC3649995.

[70] X. Zhou, P. Carbonetto, and M. Stephens. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.*, 9(2):e1003264, 2013. PMCID: PMC3567190.

[71] J. Kim, X. He, and S. Sinha. Evolution of regulatory sequences in 12 Drosophila species. *PLoS Genet.*, 5(1):e1000330, Jan 2009.

[72] J. F. Degner, A. A. Pai, R. Pique-Regi, J. B. Veyrieras, D. J. Gaffney, J. K. Pickrell, S. De Leon, K. Michelini, N. Lewellen, G. E. Crawford, M. Stephens, Y. Gilad, and J. K. Pritchard. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*, 482(7385):390–394, Feb 2012. PMCID: PMC3501342.

[73] D. Y. Lin and Z. Z. Tang. A general framework for detecting disease associations with rare variants in sequencing studies. *Am. J. Hum. Genet.*, 89(3):354–367, Sep 2011.

[74] N. Yi, N. Liu, D. Zhi, and J. Li. Hierarchical generalized linear models for multiple groups of rare and common variants: jointly estimating group and individual-variant effects. *PLoS Genet.*, 7(12):e1002382, Dec 2011.

[75] R. S. Spielman, R. E. McGinnis, and W. J. Ewens. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.*, 52(3):506–516, Mar 1993. PMCID: PMC1682161.

[76] M. A. Newton, A. Noueiry, D. Sarkar, and P. Ahlquist. Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, 5(2):155–176, Apr 2004.

[77] P. Kumar, S. Henikoff, and P. C. Ng. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 4(7):1073–1081, 2009.

[78] I. Adzhubei, D. M. Jordan, and S. R. Sunyaev. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*, Chapter 7:Unit7.20, Jan 2013. PMCID: PMC4480630.

[79] BrainSpan: Atlas of the Developing Human Brain. http://developinghumanbrain.org.

[80] The Exome Aggregation Consortium (ExAC). http://exac.broadinstitute.org/.

[81] A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, N. Abdennur, M. Adli, M. Akerman, L. Barrera, J. Antosiewicz-Bourget, T. Ballinger, M. J. Barnes, D. Bates, R. J. Bell, D. A. Bennett, K. Bianco, C. Bock, P. Boyle, J. Brinchmann, P. Caballero-Campo, R. Camahort, M. J. Carrasco-Alfonso, T. Charnecki, H. Chen, Z. Chen, J. B. Cheng, S. Cho, A. Chu, W. Y. Chung, C. Cowan, Q. Athena Deng, V. Deshpande, M. Diegel, B. Ding, T. Durham, L. Echipare, L. Edsall, D. Flowers, O. Genbacev-Krtolica, C. Gifford, S. Gillespie, E. Giste, I. A. Glass, A. Gnirke, M. Gormley, H. Gu, J. Gu, D. A. Hafler, M. J. Hangauer, M. Hariharan, M. Hatan, E. Haugen, Y. He, S. Heimfeld, S. Herlofsen, Z. Hou, R. Humbert, R. Issner, A. R. Jackson, H. Jia, P. Jiang, A. K. Johnson, T. Kadlecek, B. Kamoh, M. Kapidzic, J. Kent, A. Kim, M. Kleinewietfeld, S. Klugman, J. Krishnan, S. Kuan, T. Kutyavin, A. Y. Lee, K. Lee, J. Li, N. Li, Y. Li, K. L. Ligon, S. Lin, Y. Lin, J. Liu, Y. Liu, C. J. Luckey, Y. P. Ma, C. Maire, A. Marson, J. S. Mattick, M. Mayo, M. McMaster, H. Metsky, T. Mikkelsen, D. Miller, M. Miri, E. Mukamel, R. P. Nagarajan, F. Neri, J. Nery, T. Nguyen, H. O'Geen, S. Paithankar, T. Papayannopoulou, M. Pelizzola, P. Plettner, N. E. Propson, S. Raghuraman, B. J. Raney, A. Raubitschek, A. P. Reynolds, H. Richards, K. Riehle, P. Rinaudo, J. F. Robinson, N. B. Rockweiler, E. Rosen, E. Rynes, J. Schein, R. Sears, T. Sejnowski, A. Shafer, L. Shen, R. Shoemaker, M. Sigaroudinia, I. Slukvin, S. Stehling-Sun, R. Stewart, S. L. Subramanian, K. Suknuntha, S. Swanson, S. Tian, H. Tilden, L. Tsai, M. Urich, I. Vaughn, J. Vierstra, S. Vong, U. Wagner, H. Wang, T. Wang, Y. Wang, A. Weiss, H. Whitton, A. Wildberg, H. Witt, K. J. Won, M. Xie, X. Xing, I. Xu, Z. Xuan, Z. Ye, C. A. Yen, P. Yu, X. Zhang, X. Zhang, J. Zhao, Y. Zhou, J. Zhu, Y. Zhu, S. Ziegler, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, and M. Kellis. Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539):317–330, Feb 2015. PMCID: PMC4530010.

[82] I. Dunham, A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, C. B. Epstein, S. Frietze, J. Harrow, R. Kaul, J. Khatun, B. R. Lajoie, S. G. Landt, B. K. Lee, F. Pauli, K. R. Rosenbloom, P. Sabo, A. Safi, A. Sanyal, N. Shoresh, J. M. Simon, L. Song, N. D. Trinklein, R. C. Altshuler, E. Birney, J. B. Brown, C. Cheng, S. Djebali, X. Dong, I. Dunham, J. Ernst, T. S. Furey, M. Gerstein, B. Giardine, M. Greven, R. C. Hardison, R. S. Harris, J. Herrero, M. M. Hoffman, S. Iyer, M. Kellis, J. Khatun, P. Kheradpour, A. Kundaje, T. Lassmann, Q. Li, X. Lin, G. K. Marinov, A. Merkel, A. Mortazavi, S. C. Parker, T. E. Reddy, J. Rozowsky, F. Schlesinger, R. E. Thurman, J. Wang, L. D. Ward, T. W. Whitfield, S. P. Wilder, W. Wu, H. S. Xi, K. Y.

Yip, J. Zhuang, M. J. Pazin, R. F. Lowdon, L. A. Dillon, L. B. Adams, C. J. Kelly, J. Zhang, J. R. Wexler, E. D. Green, P. J. Good, E. A. Feingold, B. E. Bernstein, E. Birney, G. E. Crawford, J. Dekker, L. Elnitski, P. J. Farnham, M. Gerstein, M. C. Giddings, T. R. Gingeras, E. D. Green, R. Guigo, R. C. Hardison, T. J. Hubbard, M. Kellis, W. Kent, J. D. Lieb, E. H. Margulies, R. M. Myers, M. Snyder, J. A. Stamatoyannopoulos, S. A. Tenenbaum, Z. Weng, K. P. White, B. Wold, J. Khatun, Y. Yu, J. Wrobel, B. A. Risk, H. P. Gunawardena, H. C. Kuiper, C. W. Maier, L. Xie, X. Chen, M. C. Giddings, B. E. Bernstein, C. B. Epstein, N. Shoresh, J. Ernst, P. Kheradpour, T. S. Mikkelsen, S. Gillespie, A. Goren, O. Ram, X. Zhang, L. Wang, R. Issner, M. J. Coyne, T. Durham, M. Ku, T. Truong, L. D. Ward, R. C. Altshuler, M. L. Eaton, M. Kellis, S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Roder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, P. Batut, I. Bell, K. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. P. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, G. Li, O. J. Luo, E. Park, J. B. Preall, K. Presaud, P. Ribeca, B. A. Risk, D. Robyr, X. Ruan, M. Sammeth, K. S. Sandhu, L. Schaeffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, Y. Hayashizaki, J. Harrow, M. Gerstein, T. J. Hubbard, A. Reymond, S. E. Antonarakis, G. J. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigo, T. R. Gingeras, K. R. Rosenbloom, C. A. Sloan, K. Learned, V. S. Malladi, M. C. Wong, G. P. Barber, M. S. Cline, T. R. Dreszer, S. G. Heitner, D. Karolchik, W. Kent, V. M. Kirkup, L. R. Meyer, J. C. Long, M. Maddren, B. J. Raney, T. S. Furey, L. Song, L. L. Grasfeder, P. G. Giresi, B. K. Lee, A. Battenhouse, N. C. Sheffield, J. M. Simon, K. A. Showers, A. Safi, D. London, A. A. Bhinge, C. Shestak, M. R. Schaner, S. K. Kim, Z. Z. Zhang, P. A. Mieczkowski, J. O. Mieczkowska, Z. Liu, R. M. McDaniell, Y. Ni, N. U. Rashid, M. J. Kim, S. Adar, Z. Zhang, T. Wang, D. Winter, D. Keefe, E. Birney, V. R. Iyer, J. D. Lieb, G. E. Crawford, G. Li, K. S. Sandhu, M. Zheng, P. Wang, O. J. Luo, A. Shahab, M. J. Fullwood, X. Ruan, Y. Ruan, R. M. Myers, F. Pauli, B. A. Williams, J. Gertz, G. K. Marinov, T. E. Reddy, J. Vielmetter, E. Partridge, D. Trout, K. E. Varley, C. Gasper, A. Bansal, S. Pepke, P. Jain, H. Amrhein, K. M. Bowling, M. Anaya, M. K. Cross, B. King, M. A. Muratet, I. Antoshechkin, K. M. Newberry, K. McCue, A. S. Nesmith, K. I. Fisher-Aylor, B. Pusey, G. DeSalvo, S. L. Parker, S. Balasubramanian, N. S. Davis, S. K. Meadows, T. Eggleston, C. Gunter, J. Newberry, S. E. Levy, D. M. Absher, A. Mortazavi, W. H. Wong, B. Wold, M. J. Blow, A. Visel, L. A. Pennachio, L. Elnitski, E. H. Margulies, S. C. Parker, H. M. Petrykowska, A. Abyzov, B. Aken, D. Barrell, G. Barson, A. Berry, A. Bignell, V. Boychenko, G. Bussotti, J. Chrast, C. Davidson, T. Derrien, G. Despacio-Reyes, M. Diekhans, I. Ezkurdia, A. Frankish, J. Gilbert, J. M. Gonzalez, E. Griffiths, R. Harte, D. A. Hendrix, C. Howald, T. Hunt, I. Jungreis, M. Kay, E. Khurana, F. Kokocinski, J. Leng, M. F. Lin, J. Loveland, Z. Lu, D. Manthravadi, M. Mariotti, J. Mudge, G. Mukherjee, C. Notredame, B. Pei, J. M. Rodriguez, G. Saunders, A. Sboner, S. Searle, C. Sisu, C. Snow, C. Steward, A. Tanzer, E. Tapanari, M. L. Tress, M. J. van Baren, N. Walters, S. Washietl, L. Wilming, A. Zadissa, Z. Zhang, M. Brent, D. Haussler, M. Kellis, A. Valencia, M. Gerstein, A. Reymond, R. Guigo, J. Harrow, T. J. Hubbard, S. G. Landt, S. Frietze, A. Abyzov, N. Addleman, R. P. Alexander, R. K. Auerbach, S. Balasubramanian, K. Bettinger, N. Bhardwaj, A. P. Boyle, A. R. Cao, P. Cayting, A. Charos, Y. Cheng, C. Cheng, C. Eastman, G. Euskirchen, J. D. Fleming, F. Grubert, L. Habegger, M. Hariharan, A. Harmanci, S. Iyengar, V. X. Jin, K. J. Karczewski, M. Kasowski, P. Lacroute, H. Lam, N. Lamarre-Vincent, J. Leng, J. Lian, M. Lindahl-Allen, R. Min, B. Miotto, H. Monahan, Z. Moqtaderi, X. J. Mu, H. O'Geen, Z. Ouyang, D. Patacsil, B. Pei, D. Raha, L. Ramirez, B. Reed, J. Rozowsky, A. Sboner, M. Shi, C. Sisu, T. Slifer, H. Witt, L. Wu, X. Xu, K. K. Yan, X. Yang, K. Y. Yip, Z. Zhang, K. Struhl, S. M. Weissman, M. Gerstein, P. J. Farnham, M. Snyder, S. A. Tenenbaum, L. O. Penalva, F. Doyle, S. Karmakar, S. G. Landt, R. R. Bhanvadia, A. Choudhury, M. Domanus, L. Ma, J. Moran, D. Patacsil, T. Slifer, A. Victorsen, X. Yang, M. Snyder, T. Auer, L. Centanin, M. Eichenlaub, F. Gruhl, S. Heermann, B. Hoeckendorf, D. Inoue, T. Kellner, S. Kirchmaier, C. Mueller, R. Reinhardt, L. Schertel, S. Schneider, R. Sinn, B. Wittbrodt, J. Wittbrodt, Z. Weng, T. W. Whitfield, J. Wang, P. J. Collins, S. F. Aldred, N. D. Trinklein, E. C. Partridge, R. M. Myers, J. Dekker, G. Jain, B. R. Lajoie, A. Sanyal, G. Balasundaram, D. L. Bates, R. Byron, T. K. Canfield, M. J. Diegel, D. Dunn, A. K. Ebersol, T. Frum, K. Garg, E. Gist, R. Hansen, L. Boatman, E. Haugen, R. Humbert, G. Jain, A. K. Johnson, E. M. Johnson, T. V. Kutyavin, B. R. Lajoie, K. Lee, D. Lotakis, M. T. Maurano, S. J. Neph, F. V. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, E. Rynes, P. Sabo, M. E. Sanchez, R. S. Sandstrom, A. Sanyal,

A. O. Shafer, A. B. Stergachis, S. Thomas, R. E. Thurman, B. Vernot, J. Vierstra, S. Vong, H. Wang, M. A. Weaver, Y. Yan, M. Zhang, J. M. Akey, M. Bender, M. O. Dorschner, M. Groudine, M. J. MacCoss, P. Navas, G. Stamatoyannopoulos, R. Kaul, J. Dekker, J. A. Stamatoyannopoulos, I. Dunham, K. Beal, A. Brazma, P. Flicek, J. Herrero, N. Johnson, D. Keefe, M. Lukk, N. M. Luscombe, D. Sobral, J. M. Vaquerizas, S. P. Wilder, S. Batzoglou, A. Sidow, N. Hussami, S. Kyriazopoulou-Panagiotopoulou, M. W. Libbrecht, M. A. Schaub, A. Kundaje, R. C. Hardison, W. Miller, B. Giardine, R. S. Harris, W. Wu, P. J. Bickel, B. Banfai, N. P. Boley, J. B. Brown, H. Huang, Q. Li, J. J. Li, W. S. Noble, J. A. Bilmes, O. J. Buske, M. M. Hoffman, A. D. Sahu, P. V. Kharchenko, P. J. Park, D. Baker, J. Taylor, Z. Weng, S. Iyer, X. Dong, M. Greven, X. Lin, J. Wang, H. S. Xi, J. Zhuang, M. Gerstein, R. P. Alexander, S. Balasubramanian, C. Cheng, A. Harmanci, L. Lochovsky, R. Min, X. J. Mu, J. Rozowsky, K. K. Yan, K. Y. Yip, and E. Birney. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, Sep 2012. PMCID: PMC3439153.

[83] G. Macintyre, J. Bailey, I. Haviv, and A. Kowalczyk. is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics*, 26(18):i524–530, Sep 2010. PMCID: PMC2935445.

[84] A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, 15(8):1034–1050, Aug 2005. PMCID: PMC1182216.

[85] G. M. Cooper, E. A. Stone, G. Asimenos, E. D. Green, S. Batzoglou, and A. Sidow. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.*, 15(7):901–913, Jul 2005. PMCID: PMC1172034.

[86] K. G. Ardlie, D. S. Deluca, A. V. Segre, T. J. Sullivan, T. R. Young, E. T. Gelfand, C. A. Trowbridge, J. B. Maller, T. Tukiainen, M. Lek, L. D. Ward, P. Kheradpour, B. Iriarte, Y. Meng, C. D. Palmer, T. Esko, W. Winckler, J. N. Hirschhorn, M. Kellis, D. G. MacArthur, G. Getz, A. A. Shabalin, G. Li, Y. H. Zhou, A. B. Nobel, I. Rusyn, F. A. Wright, T. Lappalainen, P. G. Ferreira, H. Ongen, M. A. Rivas, A. Battle, S. Mostafavi, J. Monlong, M. Sammeth, M. Mele, F. Reverter, J. M. Goldmann, D. Koller, R. Guigo, M. I. McCarthy, E. T. Dermitzakis, E. R. Gamazon, H. K. Im, A. Konkashbaev, D. L. Nicolae, N. J. Cox, T. Flutre, X. Wen, M. Stephens, J. K. Pritchard, Z. Tu, B. Zhang, T. Huang, Q. Long, L. Lin, J. Yang, J. Zhu, J. Liu, A. Brown, B. Mestichelli, D. Tidwell, E. Lo, M. Salvatore, S. Shad, J. A. Thomas, J. T. Lonsdale, M. T. Moser, B. M. Gillard, E. Karasik, K. Ramsey, C. Choi, B. A. Foster, J. Syron, J. Fleming, H. Magazine, R. Hasz, G. D. Walters, J. P. Bridge, M. Miklos, S. Sullivan, L. K. Barker, H. M. Traino, M. Mosavel, L. A. Siminoff, D. R. Valley, D. C. Rohrer, S. D. Jewell, P. A. Branton, L. H. Sobin, M. Barcus, L. Qi, J. McLean, P. Hariharan, K. S. Um, S. Wu, D. Tabor, C. Shive, A. M. Smith, S. A. Buia, A. H. Undale, K. L. Robinson, N. Roche, K. M. Valentino, A. Britton, R. Burges, D. Bradbury, K. W. Hambright, J. Seleski, G. E. Korzeniewski, K. Erickson, Y. Marcus, J. Tejada, M. Taherian, C. Lu, M. Basile, D. C. Mash, S. Volpi, J. P. Struewing, G. F. Temple, J. Boyer, D. Colantuoni, R. Little, S. Koester, L. J. Carithers, H. M. Moore, P. Guan, C. Compton, S. J. Sawyer, J. P. Demchok, J. B. Vaught, C. A. Rabiner, N. C. Lockhart, K. G. Ardlie, G. Getz, F. A. Wright, M. Kellis, S. Volpi, and E. T. Dermitzakis. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235):648–660, May 2015. PMCID: PMC4547484.

[87] M. Kircher, D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper, and J. Shendure. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, 46(3):310–315, Mar 2014. PMCID: PMC3992975.

[88] B. Gulko, M. J. Hubisz, I. Gronau, and A. Siepel. A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat. Genet.*, 47(3):276–283, Mar 2015. PMCID: PMC4342276.

[89] H. A. Chipman, E. I. George, and R. E. McCulloch. BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, pages 266–298, 2010.

[90] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.

[91] N. Weinhold, A. Jacobsen, N. Schultz, C. Sander, and W. Lee. Genome-wide analysis of noncoding regulatory mutations in cancer. *Nat. Genet.*, 46(11):1160–1165, Nov 2014. PMCID: PMC4217527.

[92] K. Wang, M. Li, and H. Hakonarson. Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.*, 11(12):843–854, Dec 2010.

[93] Y. Jiang, Y. Han, S. Petrovski, K. Owzar, D. B. Goldstein, and A. S. Allen. Incorporating Functional Information in Tests of Excess De Novo Mutational Load. *Am. J. Hum. Genet.*, 97(2):272–283, Aug 2015.

[94] S. Akbarian, C. Liu, J. A. Knowles, F. M. Vaccarino, P. J. Farnham, G. E. Crawford, A. E. Jaffe, D. Pinto, S. Dracheva, D. H. Geschwind, J. Mill, A. C. Nairn, A. Abyzov, S. Pochareddy, S. Prabhakar, S. Weissman, P. F. Sullivan, M. W. State, Z. Weng, M. A. Peters, K. P. White, M. B. Gerstein, A. Amiri, C. Armoskus, A. E. Ashley-Koch, T. Bae, A. Beckel-Mitchener, B. P. Berman, G. A. Coetzee, G. Coppola, N. Francoeur, M. Fromer, R. Gao, K. Grennan, J. Herstein, D. H. Kavanagh, N. A. Ivanov, Y. Jiang, R. R. Kitchen, A. Kozlenkov, M. Kundakovic, M. Li, Z. Li, S. Liu, L. M. Mangravite, E. Mattei, E. Markenscoff-Papadimitriou, F. C. Navarro, N. North, L. Omberg, D. Panchision, N. Parikshak, J. Poschmann, A. J. Price, M. Purcaro, T. E. Reddy, P. Roussos, S. Schreiner, S. Scuderi, R. Sebra, M. Shibata, A. W. Shieh, M. Skarica, W. Sun, V. Swarup, A. Thomas, J. Tsuji, H. van Bakel, D. Wang, Y. Wang, K. Wang, D. M. Werling, A. J. Willsey, H. Witt, H. Won, C. C. Wong, G. A. Wray, E. Y. Wu, X. Xu, L. Yao, G. Senthil, T. Lehner, P. Sklar, and N. Sestan. The PsychENCODE project. *Nat. Neurosci.*, 18(12):1707–1712, Dec 2015.

[95] D. Warde-Farley, S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G. D. Bader, and Q. Morris. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, 38(Web Server issue):W214–220, Jul 2010.

[96] B. Zhang, C. Gaiteri, L. G. Bodea, Z. Wang, J. McElwee, A. A. Podtelezhnikov, C. Zhang, T. Xie, L. Tran, R. Dobrin, E. Fluder, B. Clurman, S. Melquist, M. Narayanan, C. Suver, H. Shah, M. Mahajan, T. Gillis, J. Mysore, M. E. MacDonald, J. R. Lamb, D. A. Bennett, C. Molony, D. J. Stone, V. Gudnason, A. J. Myers, E. E. Schadt, H. Neumann, J. Zhu, and V. Emilsson. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*, 153(3):707–720, Apr 2013. PMCID: PMC3677161.

[97] C. Colantuoni, B. K. Lipska, T. Ye, T. M. Hyde, R. Tao, J. T. Leek, E. A. Colantuoni, A. G. Elkahloun, M. M. Herman, D. R. Weinberger, and J. E. Kleinman. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*, 478(7370):519–523, Oct 2011. PMCID: PMC3510670.

[98] J. R. Gibbs, M. P. van der Brug, D. G. Hernandez, B. J. Traynor, M. A. Nalls, S. L. Lai, S. Arepalli, A. Dillman, I. P. Rafferty, J. Troncoso, R. Johnson, H. R. Zielke, L. Ferrucci, D. L. Longo, M. R. Cookson, and A. B. Singleton. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.*, 6(5):e1000952, May 2010. PMCID: PMC2869317.

[99] J. T. Glessner, K. Wang, G. Cai, O. Korvatska, C. E. Kim, S. Wood, H. Zhang, A. Estes, C. W. Brune, J. P. Bradfield, M. Imielinski, E. C. Frackelton, J. Reichert, E. L. Crawford, J. Munson, P. M. Sleiman, R. Chiavacci, K. Annaiah, K. Thomas, C. Hou, W. Glaberson, J. Flory, F. Otieno, M. Garris, L. Soorya, L. Klei, J. Piven, K. J. Meyer, E. Anagnostou, T. Sakurai, R. M. Game, D. S. Rudd, D. Zurawiecki, C. J. McDougle, L. K. Davis, J. Miller, D. J. Posey, S. Michaels, A. Kolevzon, J. M. Silverman, R. Bernier, S. E. Levy, R. T. Schultz, G. Dawson, T. Owley, W. M. McMahon, T. H. Wassink, J. A. Sweeney, J. I. Nurnberger, H. Coon, J. S. Sutcliffe, N. J. Minshew, S. F. Grant, M. Bucan, E. H. Cook, J. D. Buxbaum, B. Devlin, G. D. Schellenberg, and H. Hakonarson. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, 459(7246):569–573, May 2009. PMCID: PMC2925224.

[100] J. Y. Tzeng, P. K. Magnusson, P. F. Sullivan, and J. P. Szatkiewicz. A New Method for Detecting Associations with Rare Copy-Number Variants. *PLoS Genet.*, 11(10):e1005403, Oct 2015.

[101] B. P. Coe, K. Witherspoon, J. A. Rosenfeld, B. W. van Bon, A. T. Vulto-van Silfhout, P. Bosco, K. L. Friend, C. Baker, S. Buono, L. E. Vissers, J. H. Schuurs-Hoeijmakers, A. Hoischen, R. Pfundt, N. Krumm, G. L. Carvill, D. Li, D. Amaral, N. Brown, P. J. Lockhart, I. E. Scheffer, A. Alberti, M. Shaw, R. Pettinato, R. Tervo,

N. de Leeuw, M. R. Reijnders, B. S. Torchia, H. Peeters, B. J. O'Roak, M. Fichera, J. Y. Hehir-Kwa, J. Shendure, H. C. Mefford, E. Haan, J. Gecz, B. B. de Vries, C. Romano, and E. E. Eichler. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.*, 46(10):1063–1071, Oct 2014. PMCID: PMC4177294.

[102] X. Wen. Robust Bayesian FDR Control Using Bayes Factors, with Applications to Multi-tissue eQTL Discovery. *Statistics in Biosciences*, pages 1–22, 2016.

[103] E. E. Schadt, J. Lamb, X. Yang, J. Zhu, S. Edwards, D. Guhathakurta, S. K. Sieberts, S. Monks, M. Reitman, C. Zhang, P. Y. Lum, A. Leonardson, R. Thieringer, J. M. Metzger, L. Yang, J. Castle, H. Zhu, S. F. Kash, T. A. Drake, A. Sachs, and A. J. Lusis. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.*, 37(7):710–717, Jul 2005.

[104] L. T. Donlin, C. A. Roman, M. Adlam, A. G. Regelmann, and K. Alexandropoulos. Defective thymocyte maturation by transgenic expression of a truncated form of the T lymphocyte adapter molecule and Fyn substrate, Sin. *J. Immunol.*, 169(12):6900–6909, Dec 2002.

[105] L. T. Donlin, N. M. Danzl, C. Wanjalla, and K. Alexandropoulos. Deficiency in expression of the signaling protein Sin/Efs leads to T-lymphocyte activation and mucosal inflammation. *Mol. Cell. Biol.*, 25(24):11035–11046, Dec 2005.

[106] A. I. Chernyavsky, V. Galitovskiy, I. B. Shchepotin, and S. A. Grando. Anti-inflammatory effects of the nicotinergic peptides SLURP-1 and SLURP-2 on human intestinal epithelial cells and immunocytes. *Biomed Res Int*, 2014:609086, 2014.

[107] X. Zhu and M. Stephens. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *bioRxiv*, 2016.

[108] C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, and V. Plagnol. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, 10(5):e1004383, May 2014.

[109] A. Sekar, A. R. Bialas, H. de Rivera, A. Davis, T. R. Hammond, N. Kamitaki, K. Tooley, J. Presumey, M. Baum, V. Van Doren, G. Genovese, S. A. Rose, R. E. Handsaker, M. J. Daly, M. C. Carroll, B. Stevens, S. A. McCarroll, S. Ripke, B. M. Neale, A. Corvin, J. T. Walters, K. H. Farh, P. A. Holmans, P. Lee, B. Bulik-Sullivan, D. A. Collier, H. Huang, T. H. Pers, I. Agartz, E. Agerbo, M. Albus, M. Alexander, F. Amin, S. A. Bacanu, M. Begemann, R. A. Belliveau, J. Bene, S. E. Bergen, E. Bevilacqua, T. B. Bigdeli, D. W. Black, R. Bruggeman, N. G. Buccola, R. L. Buckner, W. Byerley, W. Cahn, G. Cai, M. J. Cairns, D. Campion, R. M. Cantor, V. J. Carr, N. Carrera, S. V. Catts, K. D. Chambert, R. C. Chan, R. Y. Chen, E. Y. Chen, W. Cheng, E. F. Cheung, S. A. Chong, C. Cloninger, D. Cohen, N. Cohen, P. Cormican, N. Craddock, B. Crespo-Facorro, J. J. Crowley, D. Curtis, M. Davidson, K. L. Davis, F. Degenhardt, J. Del Favero, L. E. DeLisi, D. Demontis, D. Dikeos, T. Dinan, S. Djurovic, G. Donohoe, E. Drapeau, J. Duan, F. Dudbridge, N. Durmishi, P. Eichhammer, J. Eriksson, V. Escott-Price, L. Essioux, A. H. Fanous, M. S. Farrell, J. Frank, L. Franke, R. Freedman, N. B. Freimer, M. Friedl, J. I. Friedman, M. Fromer, G. Genovese, L. Georgieva, E. S. Gershon, I. Giegling, P. Giusti-Rodriguez, S. Godard, J. I. Goldstein, V. Golimbet, S. Gopal, J. Gratten, L. de Haan, C. Hammer, M. L. Hamshere, M. Hansen, T. Hansen, V. Haroutunian, A. M. Hartmann, F. A. Henskens, S. Herms, J. N. Hirschhorn, P. Hoffmann, A. Hofman, M. V. Hollegaard, D. M. Hougaard, M. Ikeda, I. Joa, A. Julia, R. S. Kahn, L. Kalaydjieva, S. Karachanak-Yankova, J. Karjalainen, D. Kavanagh, M. C. Keller, B. J. Kelly, J. L. Kennedy, A. Khrunin, Y. Kim, J. Klovins, J. A. Knowles, B. Konte, V. Kucinskas, Z. A. Kucinskiene, H. Kuzelova-Ptackova, A. K. Kahler, C. Laurent, J. L. Keong, S. Lee, S. E. Legge, B. Lerer, M. Li, T. Li, K. Y. Liang, J. Lieberman, S. Limborska, C. M. Loughland, J. Lubinski, J. Lonnqvist, M. Macek, P. K. Magnusson, B. S. Maher, W. Maier, J. Mallet, S. Marsal, M. Mattheisen, M. Mattingsdal, R. W. McCarley, C. McDonald, A. M. McIntosh, S. Meier, C. J. Meijer, B. Melegh, I. Melle, R. I. Mesholam-Gately, A. Metspalu, P. T. Michie, L. Milani, V. Milanova, Y. Mokrab, D. W. Morris, O. Mors, K. C. Murphy, R. M. Murray, I. Myin-Germeys, B. Muller-Myhsok, M. Nelis, I. Nenadic, D. A. Nertney, G. Nestadt, K. K. Nicodemus, L. Nikitina-Zake, L. Nisenbaum, A. Nordin, E. O'Callaghan, C. O'Dushlaine, F. A. O'Neill, S. Y.

Oh, A. Olincy, L. Olsen, J. Van Os, C. Pantelis, G. N. Papadimitriou, S. Papiol, E. Parkhomenko, M. T. Pato, T. Paunio, M. Pejovic-Milovancevic, D. O. Perkins, O. Pietilainen, J. Pimm, A. J. Pocklington, J. Powell, A. Price, A. E. Pulver, S. M. Purcell, D. Quested, H. B. Rasmussen, A. Reichenberg, M. A. Reimers, A. L. Richards, J. L. Roffman, P. Roussos, D. M. Ruderfer, V. Salomaa, A. R. Sanders, U. Schall, C. R. Schubert, T. G. Schulze, S. G. Schwab, E. M. Scolnick, R. J. Scott, L. J. Seidman, J. Shi, E. Sigurdsson, T. Silagadze, J. M. Silverman, K. Sim, P. Slominsky, J. W. Smoller, H. C. So, C. C. Spencer, E. A. Stahl, H. Stefansson, S. Steinberg, E. Stogmann, R. E. Straub, E. Strengman, J. Strohmaier, T. Stroup, M. Subramaniam, J. Suvisaari, D. M. Svrakic, J. P. Szatkiewicz, E. Soderman, S. Thirumalai, D. Toncheva, P. A. Tooney, S. Tosato, J. Veijola, J. Waddington, D. Walsh, D. Wang, Q. Wang, B. T. Webb, M. Weiser, D. B. Wildenauer, N. M. Williams, S. Williams, S. H. Witt, A. R. Wolen, E. H. Wong, B. K. Wormley, J. Q. Wu, H. S. Xi, C. C. Zai, X. Zheng, F. Zimprich, N. R. Wray, K. Stefansson, P. M. Visscher, R. Adolfsson, O. A. Andreassen, D. H. Blackwood, E. Bramon, J. D. Buxbaum, A. D. B?rglum, S. Cichon, A. Darvasi, E. Domenici, H. Ehrenreich, T. Esko, P. V. Gejman, M. Gill, H. Gurling, C. M. Hultman, N. Iwata, A. V. Jablensky, E. G. Jonsson, K. S. Kendler, G. Kirov, J. Knight, T. Lencz, D. F. Levinson, Q. S. Li, J. Liu, A. K. Malhotra, S. A. McCarroll, A. McQuillin, J. L. Moran, P. B. Mortensen, B. J. Mowry, M. M. Nothen, R. A. Ophoff, M. J. Owen, A. Palotie, C. N. Pato, T. L. Petryshen, D. Posthuma, M. Rietschel, B. P. Riley, D. Rujescu, P. C. Sham, P. Sklar, D. St Clair, D. R. Weinberger, J. R. Wendland, T. Werge, M. J. Daly, P. F. Sullivan, and M. C. O'Donovan. Schizophrenia risk from complex variation of complement component 4. *Nature*, 530(7589):177–183, Feb 2016.

[110] K. S. Grennan, C. Chen, E. S. Gershon, and C. Liu. Molecular network analysis enhances understanding of the biology of mental disorders. *Bioessays*, 36(6):606–616, Jun 2014.

[111] M. A. Mooney, J. T. Nigg, S. K. McWeeney, and B. Wilmot. Functional and genomic context in pathway analysis of GWAS data. *Trends Genet.*, 30(9):390–400, Sep 2014.

[112] eQTL Browser - NCBI. `http://www.ncbi.nlm.nih.gov/projects/gap/eqtl/index.cgi`.

[113] seeQTL: A searchable human eQTL browser and database. `http://www.bios.unc.edu/research/genomic_software/seeQTL/`.