among SNPs in different genes can generate dispersed association signal across nearby genes, which is often misinterpreted as evidence for association of specific genes in this set. The same phenomena is relevant for PrediXcan and TWAS approaches [79], arguably amplified due to the methods used to develop the predictive model. Notably, when TWAS methods were contrasted to variant-level colocalization [81], TWAS almost always captures variant-level colocalization; however, in real data it is 100-fold more powerful. We conjecture this power traces, in part, to LD among SNPs across the GWAS locus generating false signals. We propose methods to overcome this challenge, adapting penalized regression approaches from the fairness literature.

B.2a Improving polygenic scores (PGS). PGS, also known as PRS or PGRS (polygenic risk score) when predicting disorder status, predicts phenotypes of individuals by a weighted sum of counts of their associated alleles [82]. The associated allele for each SNP and its weight are estimated from GWAS data that are independent of the samples for which PGS is to be predicted. Many methods have been proposed to construct a PGS [83–89], differing largely by which SNPs to include and how to weight them. Ni et al. [90] review and compare 10 methods popular for psychiatric genetics. Most methods outperform Pruning & Thresholding, the original PGS, and there are modest differences among most methods, with SBayesR [87] showing best overall performance in the authors hands (but see [85]). Of the methods requiring tuning, such as LDPred2 and Lassosum [86] (Lasso with summary statistics), Ni finds them to be sensitive to the tuning population. In Aim 2, we propose methods to reduce this sensitivity.

Most PGS methods use only GWAS and LD information for the score. However, as we noted earlier, functional annotations show strong enrichment for GWAS signals. For this reason, methods that select variation or weight it according to functional annotations should produce a refined score. Indeed, Marquez-Luna et al. [85] show that accounting for function, via LDPred-funct, produces a 4% relative improvement over SBayesR when fitting PGS to UK Biobank phenotypes. We propose to build on this observation. As we noted in Aim 1a, we can model annotations using gradient boosting trees to determine which annotations – and interactions of annotations – are more likely to predict causal variation and whether this varies across the myriad phenotypes evaluated by GWAS. Modeling in LDPred-funct assumes phenotypes matter, we predict that annotations effects are similar across phenotypes and will build on this conjecture if it is supported. By bringing better annotation information in PGS, we will build a more predictive PGS. In the process, we will also optimize other aspects of PGS.

B.2b Portable PGS. A PGS developed from samples of European ancestry (PGS_E) predicts the same phenotype in non-European samples poorly [91,92]. This deficit arises from a multiplicity of causes: (1) many SNPs included in the PGS_E do not have a direct effect on risk, rather they are in LD with causal variants and LD patterns vary across populations; (2) causal variants have somewhat different effects across populations; (3) the genetic architecture of the trait differs somewhat among populations. If (1) causes the greatest lack-of-portability, as we expect, then the solution lies in fixing the PGS_E, as proposed in Aim 2.2a, and is supported by recent studies [46, 93, 94]. A simple solution for (2) is to combine non-European PRS (PRS_N) with a PRS_E in an optimal way. Marquez-Luna et al. [95] develop such an approach, estimating optimal weights for a trans-ethnic pruning and thresholding PGS by training on both populations. Here we propose a solution to (2), building on Lassosum and taking two approaches, one in the spirit of Marquez-Luna et al. and another joint optimization of Lassosum over populations. Because some populations are of admixed ancestry, we will extend these ideas using local ancestry segments [96–98] to obtain portable PGS for a variety of population histories. We will introduce measures and concepts of fairness, a critical feature of PGS non-portability [99], to the human genetic literature. Different genetic architectures (3) creates a difficult hurdle, especially if the populations differ substantially by environmental effects. From a purely genetic perspective, solutions for (1) and (2) are still the best approach for a portable PGS, while incorporating non-genetic terms would be invaluable for prediction.

B.3 Models for rare non-coding variants. To relate clinical outcome to rare coding or copy number variants (CNVs), clinical geneticists typically used a Fisher Exact Test. Recognizing that the FET sacrifices power because it ignores critical information about genes/genomic regions, their mutation rates, and the properties of the mutations, in 2011 we built a more powerful model for CNV association [100]. We then extended the idea to variation in coding sequence [101, 102], and next introduced the Transmission And De novo Association framework [103], which modeled sequence variation of three inheritance types: de novo, inherited, or from case-control samples. The current version of TADA integrates all classes of coding-region variation, including CNVs [30]. Using TADA, we have identified hundreds of genes associated with autism and other neurodevelopmental disorders [20,30,104] and have also illuminated its neurobiology [30,105,106]. Key to TADA's power is prior information about the nature of genes (e.g., level of conservation) and genetic variants (e.g., protein truncating (PTV), conserved missense, silent), as well as gene-level mutation rates. The prior information is critical and revealing. For instance, of sequence variation, de novo PTVs found in conserved genes have the largest effect on risk; yet, PTVs transmitted from unaffected parents to their affected offspring confer far less risk, even when they fall in the

CommonMind and PsychEncode Consortia [21, 51, 52, 74, 148], which have genetic and gene expression data from brain tissue. (3) A new resource (R01MH125235) – partially overlapping with CMC resources – but also producing proteomics data (levels of protein expression, phosphorylation, glycosylation, and ubiquitination) on 400 brain samples. These samples also are, or will be, characterized for gene expression and whole genome sequence (please see Matt MacDonald's letter of support). (4) BrainVar, a PsychEncode project that will characterize ≈ 100 human brains, performing temporal, regional, and cell-type-specific transcriptome profiling (snRNA-seq) of the developing human brain. The DNA samples will also be characterized for ATAC-seq peaks and whole genome sequence (please see Stephan Sander's letter of support). And (5), other relevant, new PsychEncode projects. Of these, (2) involves Devlin as an mPI (with Matt MacDonald and Jon Trinidad), thus the data are freely available for validation of methods proposed here and for collaborative research between our groups. (Remark: While colocalization is a goal of R01MH125235, only Ensemble approaches, which combine existing methods, are proposed therein. Thus, there is no methodological overlap with the research proposed here, although we do draw inspiration from it.) For (4) both Devlin and Roeder are involved as collaborators for the BrainVar study (Stephan Sanders and Nenad Sestan, mPIs). Hence the data will be available for validation of methods. (Again, there is no overlap regarding methods development of the BrainVar project and what we propose here.)

AIM 2: *Improving polygenic scores and adapting them for diverse ancestries*

Background *The objective of PGS is to predict a phenotype by a linear model of SNP genotypes. If we knew all of the causal SNPs in a population, the problem would simplify to weighting the SNPs properly. Due to LD and low power, however, we don't know the causal SNPs. Moreover, even in a GWAS locus, LD among SNPs obscures the causal variant(s), this correlation tends to be blocky, and the size of these blocks varies by ancestry. For these reasons, PGS tend to include many "linked or tag" SNPs, rather than causal variants. And, for this reason, predictors built from Eurocentric samples do not successfully port to other ancestries. To improve a PGS, and make it more portable, we should include functional SNPs rather than tag SNPs. Several efforts in the literature have had some success in this domain [85], but there is room for further improvement. Furthermore, we recognize that some populations are of admixed ancestry. For this reason, we will extend the ideas outlined below to account for it by using local ancestry segments [96–98].*

Aim 2a: Building a more predictive PGS *There are several methods in the literature for injecting annotations into PGS. Our primary contribution will be in assimilating the information in a powerful and interpretable form. We will investigate available methods for how best to utilize this information in a PGS form. We will apply our more elaborate modeling of functionality, as described in Aim 1a, to identify SNPs that are more likely functional and thereby improve the PGS. However, Aim 1a only targets GWAS loci, whereas a good PGS typically includes a wider set of SNPs. In [7] we lay out an approach using AdaPT for selective inference and XGBoost for model fitting and by which we identify likely causal SNPs by virtue of their annotation and GWAS p-values – even when the p-values are far from GWAS-significant (i.e., $> 5 \times 10^{-8}$). We will explore this approach for selecting SNPs for the PGS. Furthermore, as noted in Aim 1, BSLMM already provides a relatively sparse solution and itself produces an improved PGS [139]. We will seek to hybridize the features of BSLMM and the AdaPT/XGBoost approach to introduce an informative prior and hence produce a refined PGS. (Remark: BSLMM is closely related to SuSiE and therefore we believe this is possible.) Finally, we will explore CVC as a possible replacement for BSLMM and compare results.*

Aim 2b: Building a more portable PGS. *Here we seek a PGS with maximum portability across ancestries. Typically we have available large training data sets for the majority population, usually Euro-centric (EC), and only modest sized samples for the other population, not EC (nEC). Signals are weak, so large amounts of data are essential for success in the training stage, and yet it will take time to collect such samples from populations whose ancestries are nEC. Ideally we could learn from all the data, but avoid the biases inherent in using primarily the EC sample.*

To formulate and validate a PGS model, ideally a large GWAS is available for fitting the model (training), a smaller independent dataset is available from which to choose the tuning parameters (testing) and finally an independent dataset can be utilized for validation. However, commonly no testing data are available, especially for the nEC population. Moreover, for many phenotypes, only summary statistics are available, precluding the option of using cross-validation on the training data. These practical issues create statistical challenges.

We focus on a data structure with $Y \in \{0,1\}^n$ indicating case-control status, $X \in \mathbb{R}^{n \times (p+1)}$ an intercept and centered variables indicating genotype for p SNPs. (Remark: Methods pertain to continuous phenotypes also.) We want to learn a linear model that predicts Y from X using the lasso estimator, which can be expressed as $\widehat{\beta}_\lambda = \arg\min 2R^T\beta + \beta^T C\beta + \lambda\|\beta\|_1$ where $R = \frac{1}{n}X^TY$ and $C = \frac{1}{n}X^TX$ are the empirical correlation between X and Y, and the empirical covariance of X (i.e., LD), respectively. The Lassosum estimator [86], replaces C with a regularized version of the same obtained from a library such as 1000 genomes. The performance of the

lasso estimator is sensitive to the choice of λ [90], hence reliable methods for selection of the tuning parameter are needed.

When sampling from two ancestries, we anticipate that the signal, β, is similar across ancestries, but that the LD pattern varies. Our proposed method, which we call Joint-Lasso, considers a linear combination of loss functions from the two ancestries, which simplifies to a convenient form. Let $C = \gamma C_1 + (1-\gamma)C_2$ and $R = \gamma R_1 + (1-\gamma R_2)$, where C_j, R_j are the sample correlations from population $j \in \{1,2\}$ and substitute these into the lasso formulation. In this setting, we ultimately need to choose three tuning parameters: the mixing parameter γ and population dependent λ's to obtain a small risk within each ancestry group. It is challenging to chose the tuning parameters because we only have available summary statistics from a single source. This precludes traditional approaches to cross-validation for training and testing. We focus discussion on λ, but the same ideas apply to choice of γ. We note that all results below can be simplified for the one ancestry problem.

We explored performance of Joint-Lasso and tuning parameters via simulations. Summary statistics were generated for a training and testing sample. For the training data, we using EC and African 1000 Genomes data to generate genotypes with realistic LD structure, then selected p SNPs as causal variants to achieve a certain total h^2_T, which could be the same or different between populations. Causal variants contributed equally to h^2_T, but the effects varied as a function of population-based allele frequency. Sample sizes varied, but usually were much larger for the EC sample. Tests of association (GWAS) were then performed to yield summary statistics. We found that Joint-Lasso produces a PGS that performs well for both ancestries and is superior to existing methods targeting portability. Importantly, if we used extreme PGS values from lassosum as a simple classification rule, false positives were far more likely for African than for EC ancestry. For Joint-lasso and the same classification rule, however, the false classification rate was substantially reduced. In addition, from these simulations we obtained preliminary evidence that the optimal choice of λ varies by population, largely due to the information content for the PGS, but that γ is a function of the relative sample sizes and to a far lesser extent, heritability. We will develop these ideas further.

Selecting tuning parameters based on summary statistics *The tuning of λ aims at minimizing each predictive risk: $\mathrm{Risk}_j(\widehat{\beta}_\lambda) = \mathbb{E}[(Y^{(j)}_{\mathrm{new}} - (X^{(j)})^T\widehat{\beta}_\lambda)^2|\widehat{\beta}_\lambda]$ where $Y^{(j)}_{\mathrm{new}}$ is an independent draw from the jth population. We assume fixed $X^{(j)}$ and the expectation is over $Y^{(j)}$ and $Y^{(j)}_{\mathrm{new}}$. Based on a result of Efron [149], the predictive risk can be related to the in-sample predictive risk as follows: $\mathrm{Risk}_j = \|Y^{(j)} - X^{(j)}\widehat{\beta}_\lambda\|_2^2 + 2\sum_{i=1}^{n_j}\mathrm{Cov}(\widehat{\mu}^{(j)}_i, Y^{(j)}_i)$, where $\widehat{\mu}^{(j)}_i = (X^{(j)}_i)^T\widehat{\beta}$ is the fitted mean value for $\mathbb{E}(Y^{(j)}_i|X^{(j)}_i)$. The term $\mathrm{Cov}(\widehat{\mu}_i, y_i)$ is called the "optimism" (also known as "covariance penalty"), which quantifies the difference between the actual predictive risk and the in-sample predictive risk. (It is easy to check that the in-sample predictive risk can be computed using only the summary statistics.)*

Let $(Y^{(1)*}_1, ..., Y^{(1)*}_{n_1})$, $(Y^{(2)*}_1, ..., Y^{(2)*}_{n_2})$ be a bootstrap sample, then we can approximate the optimism term by $\widehat{\mathrm{Cov}}(\widehat{\mu}^{(j)}_i, Y^{(j)}_i) = \mathrm{Cov}_*(\widehat{\mu}^{(j)*}_i, y^{(j)*}_i)$, where $\mu^{(j)*}_i = (X^{(j)}_i)^T\widehat{\beta}^*$, and $\widehat{\beta}^*$ is the bootstrap version of $\widehat{\beta}$ using the bootstrap sample. The bootstrap sample is generated by $y^{(j)*}_i = \widehat{\mu}^{(j)}_{i,0} + \epsilon^{(j)*}_i$ where $\epsilon^{(j)*}_i$ is centered Bernoulli noise such that $\mathbb{E}_*\epsilon^{(j)*}_i = 0$ and $\mathbb{E}_*(\epsilon^{(j)*}_i)^2 = \widehat{\mu}^{(j)}_{i,0}(1-\widehat{\mu}^{(j)}_{i,0})$. Here $\widehat{\mu}_0$ is a "preliminary estimate", which is expected to be fairly accurate although not optimal. Such an estimate can usually be obtained using a relatively small λ. The challenge is to approximate the bootstrap procedure when only summary statistics are available.

Now $\sum_{i=1}^n \mathrm{Cov}_*(\widehat{\mu}^{(j)*}_i, Y^{(j)*}_i) = \mathbb{E}_*(\widehat{\beta}^*)^T(X^{(j)})^T\epsilon^{(j)*}$. And to obtain $\widehat{\beta}^*$, we will need the bootstrap versions of R_1 and R_2, where $R^*_j = (X^{(j)})^TY^{(j)*} = (X^{(j)})^T(X^{(j)}\widehat{\mu}_0 + \epsilon^{(j)*}) = n_jC_j\widehat{\mu}_0 + (X^{(j)})^T\epsilon^{(j)*}$. Therefore, to obtain the optimism estimate, we only need to generate $(X^{(j)})^T\epsilon^{(j)*}$ for $j = 1, 2$, but X is not recorded. To achieve our goal, let $X^{(j)} = U_jD_jV_j^T$ be the singular value decomposition of $X^{(j)}$. If we only have the summary statistic C_j, then U_j is not accessible but (V_j, D_j) is. Because U_j is full rank orthonormal, we only need to generate $\tilde{\epsilon}^*$, where $\tilde{\epsilon}^* = U^T\epsilon^*$. The first and second moments of $\tilde{\epsilon}^{(j)*}$ are: $\mathbb{E}(\epsilon^{(j)*}) = 0$ and $\mathbb{E}\epsilon^{(j)*}(\epsilon^{(j)*})^T = \mathrm{diag}(\widehat{\mu}^{(j)}_{1,0} - (\widehat{\mu}^{(j)}_{1,0})^2, ..., \widehat{\mu}^{(j)}_{n,0} - (\widehat{\mu}^{(j)}_{n,0})^2)$. We can use the following a "partially-second-order" Gaussian approximation $\tilde{\epsilon}^{(j)*} \sim N(0, \tau_j^2 I_{n_j})$ where $\tau_j^2 = \widehat{\beta}_{1,0} + \widehat{\beta}_0^TC_j\widehat{\beta}_0$. Here $\widehat{\beta}_{1,0}$ is the first coordinate (i.e., intercept) of the preliminary estimate $\widehat{\beta}_0$, and we used the fact that the other columns in $X^{(j)}$ are centered and hence sum to 0.

Finally, given summary statistics R_j, C_j for $j = 1, 2$, corresponding sample sizes n_j, preliminary estimate $\widehat{\beta}_0$ and bootstrap sample size B, we implement the following procedure: for $b = 1, ..., B$, $j = 1, 2$, generate $\tilde{\epsilon}^{j(*)}$; emulate $(X^{(j)})^T\epsilon^{(j)*}$ by $V_jD_j\tilde{\epsilon}^{(j)*}$ where $V_jD_j^2V_j^T$ is the SVD of nC_j; and compute $\widehat{\beta}^*$ using (R^*_1, C_1, R^*_2, C_2). The covariance penalty is approximated by the average of $(\widehat{\beta}^*)^T(X^{(j)})^T\epsilon^{(j)*}$ over the bootstrap repetitions.

A and B, there are three possible intersections: $A = 1$ or $B = 1$; $A = 1$ & $B = 1$. The considerable overlap between annotation intersections leads to test statistics with substantial levels of correlation, which confounds the interpretability and error rate guarantees of multiple testing procedures. To overcome this challenge we use an agglomerative procedure that was implemented in the context of gene-level testing [11] to cluster highly correlated intersections together for testing. Ultimately our tests can discover clusters of significant annotation intersections. We conjecture these results can be interpreted downstream using new data blurring techniques to enable us to estimate valid post-selection confidence intervals for the effect sizes of the CWAS annotation intersections [150].

Using realistic simulations, our preliminary results show that the ideas described above will work in practice, but we still lack power. To enhance performance we will implement selective inference approaches [191], which we have successfully implemented in the context of SNP and gene-based tests [7, 11]. In this setting, a new twist is needed because the side-information of annotations also defines the categories. Luckily, the correlation structure of intersections can be exploited, allowing us to pool shared information across their intersections as metadata for guiding our multiple testing correction. Specifically, we leverage the overlapping annotation metadata using XGBoost in the AdaPT framework to up-weight hypotheses that are more likely to be non-null. (Due to the discreteness of performing Poisson enrichment tests, we rely on innovations in selective inference masking functions [192, 193] to ensure we retain power in this setting.) Preliminary investigations with simulated data indicate that pooling annotation information in this manner can lead to substantially improved power [194].

Challenges *(1) To evaluating meta-learning, we can exploit the abundance of data and knowledge of variation in the exome. We know MPC scores can roughly separate de novo missense variants into benign and risk variants; can meta-learning match or exceed this success by the evaluation of transmitted missense variation, then applying what we learned to de novo missense variation? (2) We describe our meta-learning approach for parent-offspring trios, from whom we can infer transmitted and de novo variants. Can this type of analysis be modified to incorporate large data sets with standing variation, such as TOPMed, by using a mutational model to predict which annotations and intersections of annotations show a dearth of variation (akin to derivation of LOEUF)? (3) The objectives of meta-learning and selective inference have overlapping goals. We will investigate whether the model learned with meta-learning can bolster the power of CWAS and selective inference by refining and potentially reducing the number of tests performed. (4) We will evaluate whether and how our findings from meta-learning can be used to improve various LR methods for testing non-coding variants. (5) We will pair our meta-learning results with the TADA-A [129] algorithm.*

Data for Aim 3. *Resources for WGS data are expanding and we expect substantial datasets in the near future. For instance, Alzheimer's Disease Sequencing Project [195] recently released WGS data from 16,906 samples, available through NIAGADS, and they anticipate 20,000 additional samples soon (please see Li-San Wang's letter of support). TOPMed Freeze 9 has WGS data from 206,000 individuals, with samples from CCDG, 158,470 TOPMed samples and 2,504 1000 Genomes samples. For mental disorders, Devlin and Roeder are mPIs of the ASC-SSC Whole Genome Consortium (along with Mike Talkowski, Stephan Sanders, and Joe Dougherty; (please see Mike Talkowski's and Stephan Sanders' letters of support)), formed for the analysis of WGS data from ASD families. We anticipate aggregating WGS data from 8,626 ASD probands within 8,189 families this year. In terms of other WGS data for mental health, the Whole Genome Sequencing for Psychiatric Disorders Consortium is also aggregating data. As they note in their publication, "The Whole Genome Sequencing for Psychiatric Disorders Consortium will integrate data for 18,000 individuals with psychiatric disorders, beginning with autism spectrum disorder, schizophrenia, bipolar disorder, and major depressive disorder, along with over 150,000 controls." [196] A data set of special relevance to ASD is congenital heart disease (CHD). Like ASD, CHD is uncommon, occurring in roughly 1% of births. Rare genetic variants, especially de novo variants that damage genes, occur at an elevated rate in CHD individuals. Furthermore, CHD is associated with neurodevelopmental disorders and must be under negative selection. Notably, WGS data are currently available from 763 CHD probands [197] and their families and data from 1812 should be available soon [198]. These data should be an excellent training set from which to learn about important non-coding variation; in turn, one could say the same for the ASD data, it would make an excellent training set of CHD. Clearly both would benefit from the methods we propose here.*

Timeline *We do not anticipate working on the Aims in the order they appear, we will be working on them simultaneously. All are timely and the field is competitive. We expect releasing manuscripts from all three Aims during the first year of the project, with follow-up and additional studies thereafter. During Year 4 we anticipate working on new topics, as well as completing the Aims described herein, in anticipation of a renewal application.*

Sex and other relevant biological variables *Our methods apply regardless of sex.*

References

[1] *L. Zhu, J. Lei, B. Devlin, and K. Roeder. A UNIFIED STATISTICAL FRAMEWORK FOR SINGLE CELL AND BULK RNA SEQUENCING DATA.* Ann Appl Stat*, 12(1):609–632, Mar 2018. PMCID: PMC6114100.*

[2] *F. Liu, D. Choi, L. Xie, and K. Roeder. Global spectral clustering in dynamic networks.* Proc Natl Acad Sci U S A*, 115(5):927–932, 01 2018. PMCID: PMC5798376.*

[3] *L. Zhu, J. Lei, L. Klei, B. Devlin, and K. Roeder. Semisoft clustering of single-cell data.* Proc Natl Acad Sci U S A*, 116(2):466–471, 01 2019. PMCID: PMC6329952.*

[4] *J.-Y. An, K. Lin, L. Zhu, D. M. Werling, S. Dong, H. Brand, H. Z. Wang, X. Zhao, G. B. Schwartz, R. L. Collins, B. B. Currall, C. Dastmalchi, J. Dea, C. Duhn, M. C. Gilson, L. Klei, L. Liang, E. Markenscoff-Papadimitriou, S. Pochareddy, N. Ahituv, J. D. Buxbaum, H. Coon, M. J. Daly, Y. S. Kim, G. T. Marth, B. M. Neale, A. R. Quinlan, J. L. Rubenstein, N. Sestan, M. W. State, A. J. Willsey, M. E. Talkowski, B. Devlin, K. Roeder, and S. J. Sanders. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder.* Science*, 362(6420), 12 2018. PMCID: PMC6432922.*

[5] *M. E. Hauberg, J. F. Fullard, L. Zhu, A. T. Cohain, C. Giambartolomei, R. Misir, S. Reach, J. S. Johnson, M. Wang, M. Mattheisen, A. D. Børglum, B. Zhang, S. K. Sieberts, M. A. Peters, E. Domenici, E. E. Schadt, B. Devlin, P. Sklar, K. Roeder, P. Roussos, and CommonMind Consortium. Differential activity of transcribed enhancers in the prefrontal cortex of 537 cases with schizophrenia and controls.* Mol Psychiatry*, 24(11):1685–1695, 11 2019. PMCID: PMC6222027.*

[6] *J. Wang, B. Devlin, and K. Roeder. Using multiple measurements of tissue to estimate subject- and cell-type-specific gene expression.* Bioinformatics*, 36(3):782–788, 02 2020. PMCID: PMC7523682.*

[7] *R. Yurko, M. G'Sell, K. Roeder, and B. Devlin. A selective inference approach for false discovery rate control using multiomics covariates yields insights into disease risk.* Proc Natl Acad Sci U S A*, 117(26):15028–15035, 06 2020. PMCID: PMC7334489.*

[8] *S. Chen, J. Wang, E. Cicek, K. Roeder, H. Yu, and B. Devlin. De novo missense variants disrupting protein-protein interactions affect risk for autism through gene co-expression and protein networks in neuronal cell types.* Mol Autism*, 11(1):76, 10 2020. PMCID: PMC7545940.*

[9] *J. Wang, K. Roeder, and B. Devlin. Bayesian estimation of cell type-specific gene expression with prior derived from single-cell data.* Genome Res*, 31(10):1807–1818, 10 2021. PMCID: PMC8494232.*

[10] *Y. Qiu, J. Lei, and K. Roeder. Gradient-based Sparse Principal Component Analysis with Extensions to Online Learning.* arXiv preprint arXiv:1911.08048*, 2020.*

[11] *R. Yurko, K. Roeder, B. Devlin, and M. G'Sell. H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives.* Ann Hum Genet*, 85(3-4):97–100, 05 2021. PMID: 33372276.*

[12] *X. Wang, D. Choi, and K. Roeder. Constructing local cell-specific networks from single-cell data.* Proc Natl Acad Sci U S A*, 118(51), 12 2021. PMCID: PMC8713783.*

[13] *M. Jalbrzikowski, F. Liu, W. Foran, K. Roeder, B. Devlin, and B. Luna. Resting-State Functional Network Organization Is Stable Across Adolescent Development for Typical and Psychosis Spectrum Youth.* Schizophr Bull*, 46(2):395–407, 02 2020. PMCID: PMC7442350.*

[14] *M. Jalbrzikowski, F. Liu, W. Foran, L. Klei, F. J. Calabro, K. Roeder, B. Devlin, and B. Luna. Functional connectome fingerprinting accuracy in youths and adults is similar when examined on the same day and 1.5-years apart.* Hum Brain Mapp*, 41(15):4187–4199, 10 2020. PMCID: PMC7502841.*

[15] *K. Z. Lin, H. Liu, and K. Roeder. Covariance-based sample selection for heterogeneous data: Applications to gene expression and autism risk gene detection.* J Am Stat Assoc*, 116(533):54–67, 2021. PMCID: PMC7958652.*

[16] *Y. Qiu, J. Wang, J. Lei, and K. Roeder. Identification of cell-type-specific marker genes from co-expression patterns in tissue samples.* Bioinformatics*, Apr 2021. PMCID: PMC8504631.*

[17] *M. Peng, B. Wamsley, A. G. Elkins, D. H. Geschwind, Y. Wei, and K. Roeder. Cell type hierarchy reconstruction via reconciliation of multi-resolution cluster tree.* Nucleic Acids Res*, 49(16):e91, 09 2021. PMCID: PMC8450107.*

[18] *M. Peng, Y. Li, B. Wamsley, Y. Wei, and K. Roeder. Integration and transfer learning of single-cell transcriptomes via cFIT.* Proc Natl Acad Sci U S A*, 118(10), 03 2021. PMCID: PMC7958425.*

[19] *N. L. Oliveira, J. Lei, and R. J. Tibshirani. Unbiased Risk Estimation in the Normal Means Problem via Coupled Bootstrap Techniques.* arXiv preprint arXiv:2111.09447*, 2021.*

[20] *F. K. Satterstrom, J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J.-Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias, Autism Sequencing Consortium, iPSYCH-Broad Consortium, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, and J. D. Buxbaum. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism.* Cell*, 180(3):568–584.e23, 02 2020. PMCID: PMC7250485.*

[21] *G. E. Hoffman, J. Bendl, G. Voloudakis, K. S. Montgomery, L. Sloofman, Y.-C. Wang, H. R. Shah, M. E. Hauberg, J. S. Johnson, K. Girdhar, L. Song, J. F. Fullard, R. Kramer, C.-G. Hahn, R. Gur, S. Marenco, B. K. Lipska, D. A. Lewis, V. Haroutunian, S. Hemby, P. Sullivan, S. Akbarian, A. Chess, J. D. Buxbaum, G. E. Crawford, E. Domenici, B. Devlin, S. K. Sieberts, M. A. Peters, and P. Roussos. CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder.* Sci Data*, 6(1):180, 09 2019. PMCID: PMC6760149.*

[22] *M. S. Breen, A. Dobbyn, Q. Li, P. Roussos, G. E. Hoffman, E. Stahl, A. Chess, P. Sklar, J. B. Li, B. Devlin, J. D. Buxbaum, and CommonMind Consortium. Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia.* Nat Neurosci*, 22(9):1402–1412, 09 2019. PMCID: PMC6791127.*

[23] *L. M. Huckins, A. Dobbyn, D. M. Ruderfer, G. Hoffman, W. Wang, A. F. Pardiñas, V. M. Rajagopal, T. D. Als, H. T Nguyen, K. Girdhar, J. Boocock, P. Roussos, M. Fromer, R. Kramer, E. Domenici, E. R. Gamazon, S. Purcell, CommonMind Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, iPSYCH-GEMS Schizophrenia Working Group, D. Demontis, A. D. Børglum, J. T. R. Walters, M. C. O'Donovan, P. Sullivan, M. J. Owen, B. Devlin, S. K. Sieberts, N. J. Cox, H. K. Im, P. Sklar, and E. A. Stahl. Gene expression imputation across multiple brain regions provides insights into schizophrenia risk.* Nat Genet*, 51(4):659–674, 04 2019. PMCID: PMC7034316.*

[24] *D. M. Werling, S. Pochareddy, J. Choi, J.-Y. An, B. Sheppard, M. Peng, Z. Li, C. Dastmalchi, G. Santpere, A. M. M. Sousa, A. T. N. Tebbenkamp, N. Kaur, F. O. Gulden, M. S. Breen, L. Liang, M. C. Gilson, X. Zhao, S. Dong, L. Klei, A. E. Cicek, J. D. Buxbaum, H. Adle-Biassette, J.-L. Thomas, K. A. Aldinger, D. R. O'Day, I. A. Glass, N. A. Zaitlen, M. E. Talkowski, K. Roeder, M. W. State, B. Devlin, S. J. Sanders, and N. Sestan. Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic Coordination in the Developing Human Prefrontal Cortex.* Cell Rep*, 31(1):107489, 04 2020. PMCID: PMC7295160.*

[25] *B. Mahjani, L. Klei, C. M. Hultman, H. Larsson, B. Devlin, J. D. Buxbaum, S. Sandin, and D. E. Grice. Maternal Effects as Causes of Risk for Obsessive-Compulsive Disorder.* Biol Psychiatry*, 87(12):1045–1051, 06 2020. PMCID: PMC8023336.*

[26] *F. L. Wang, S. L. Pedersen, B. Devlin, E. M. Gnagy, W. E. Pelham, Jr, and B. S. G. Molina. Heterogeneous Trajectories of Problematic Alcohol Use, Depressive Symptoms, and their Co-Occurrence in Young Adults with and without Childhood ADHD.* J Abnorm Child Psychol*, 48(10):1265–1277, 10 2020. PMCID: PMC7470627.*

[27] *B. Mahjani, S. De Rubeis, C. Gustavsson Mahjani, M. Mulhern, X. Xu, L. Klei, F. K. Satterstrom, J. Fu, M. E. Talkowski, A. Reichenberg, S. Sandin, C. M. Hultman, D. E. Grice, K. Roeder, B. Devlin, and J. D. Buxbaum. Prevalence and phenotypic impact of rare potentially damaging variants in autism spectrum disorder.* Mol Autism*, 12(1):65, 10 2021. PMCID: PMC8495954.*

[28] *L. Klei, L. L. McClain, B. Mahjani, K. Panayidou, S. De Rubeis, A.-C. S. Grahnat, G. Karlsson, Y. Lu, N. Melhem, X. Xu, A. Reichenberg, S. Sandin, C. M. Hultman, J. D. Buxbaum, K. Roeder, and B. Devlin. How rare and common risk variation jointly affect liability for autism spectrum disorder.* Mol Autism*, 12(1):66, 10 2021. PMCID: PMC8495987.*

[29] *B. Mahjani, L. Klei, M. Mattheisen, M. W. Halvorsen, A. Reichenberg, K. Roeder,* N. L. Pedersen, J. Boberg, *E. de Schipper, C. M. Bulik, M. Landén, B. Fundín, D. Mataix-Cols, S. Sandin, C. M. Hultman, J. J. Crowley, J. D. Buxbaum, C. Rück, B. Devlin, and D. E. Grice. The Genetic Architecture of Obsessive-Compulsive Disorder: Contribution of Liability to OCD From Alleles Across the Frequency Spectrum.* Am J Psychiatry*, 179(3):216–225, Mar 2022. PMCID: PMC8897260.*

[30] *J. M. Fu, F. K. Satterstrom, M. Peng, H. Brand, R. L. Collins, S. Dong, L. Klei, C. R. Stevens, C. Cusick, M. Babadi, E. Banks, B. Collins, S. Dodge, S. B. Gabriel, L. Gauthier, S. K. Lee, L. Liang, A. Ljungdahl, B. Mahjani, L. Sloofman, A. Smirnov, M. Barbosa, A. Brusco, B. H. Chung, M. L. Cuccaro, E. Domenici, G. B. Ferrero, J. J. Gargus, G. E. Herman, I. Hertz-Picciotto, P. Maciel, D. S. Manoach, M. R. Passos-Bueno, A. M. Persico, A. Renieri, F. Tassone, E. Trabetti, G. Campos, M. C. Chan, C. Fallerini, E. Giorgio, A. C. Girard, E. Hansen-Kiss, S. L. Lee, C. Lintas, Y. Ludena, R. Nguyen, L. Pavinato, M. Pericak-Vance, I. Pessah, E. Riberi, R. Schmidt, M. Smith, C. I. Souza, S. Trajkova, J. Y. Wang, M. H. Yu, T. A. S. C. (ASC), B. I. C. for Common Disease Genomics (Broad-CCDG), iPSYCH BROAD Consortium, D. J. Cutler, S. De Rubeis, J. D. Buxbaum, M. J. Daly, B. Devlin, K. Roeder, S. J. Sanders, and M. E. Talkowski. Rare coding variation illuminates the allelic architecture, risk genes, cellular expression patterns, and phenotypic context of autism.* medRxiv, (Nat Genet, in press)*, 2021.*

[31] *M. Cai, M. Yue, T. Chen, J. Liu, E. Forno, X. Lu, T. Billiar, J. Celedón, C. McKennan, W. Chen, and J. Wang. Robust and accurate estimation of cellular fraction from tissue omics data via ensemble deconvolution.* Bioinformatics*, Apr 2022. PMID: 35438146.*

[32] *D. J. Schaid, W. Chen, and N. B. Larson. From genome-wide associations to candidate causal variants by statistical fine-mapping.* Nat Rev Genet*, 19(8):491–504, 08 2018. PMCID: PMC6050137.*

[33] *Wellcome Trust Case Control Consortium, J. B. Maller, G. McVean, J. Byrnes, D. Vukcevic, K. Palin, Z. Su, J. M. M. Howson, A. Auton, S. Myers, A. Morris, M. Pirinen, M. A. Brown, P. R. Burton, M. J. Caulfield, A. Compston, M. Farrall, A. S. Hall, A. T. Hattersley, A. V. S. Hill, C. G. Mathew, M. Pembrey, J. Satsangi, M. R. Stratton, J. Worthington, N. Craddock, M. Hurles, W. Ouwehand, M. Parkes, N. Rahman, A. Duncanson, J. A. Todd, D. P. Kwiatkowski, N. J. Samani, S. C. L. Gough, M. I. McCarthy, P. Deloukas, and P. Donnelly. Bayesian refinement of association signals for 14 loci in 3 common diseases.* Nat Genet*, 44(12):1294–301, Dec 2012. PMCID: PMC3791416.*

[34] *F. Hormozdiari, E. Kostem, E. Y. Kang, B. Pasaniuc, and E. Eskin. Identifying causal variants at loci with multiple signals of association.* Genetics*, 198(2):497–508, Oct 2014. PMCID: PMC4196608.*

[35] *W. Chen, B. R. Larrabee, I. G. Ovsyannikova, R. B. Kennedy, I. H. Haralambieva, G. A. Poland, and D. J. Schaid. Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics.* Genetics*, 200(3):719–36, Jul 2015. PMCID: PMC4512539.*

[36] *H. Huang, M. Fang, L. Jostins, M. Umićević Mirkov, G. Boucher, C. A. Anderson, V. Andersen, I. Cleynen, A. Cortes, F. Crins, M. D'Amato, V. Deffontaine, J. Dmitrieva, E. Docampo, M. Elansary, K. K.-H. Farh, A. Franke, A.-S. Gori, P. Goyette, J. Halfvarson, T. Haritunians, J. Knight, I. C. Lawrance, C. W. Lees, E. Louis, R. Mariman, T. Meuwissen, M. Mni, Y. Momozawa, M. Parkes, S. L. Spain, E. Théâtre, G. Trynka, J. Satsangi, S. van Sommeren, S. Vermeire, R. J. Xavier, International Inflammatory Bowel Disease Genetics Consortium, R. K. Weersma, R. H. Duerr, C. G. Mathew, J. D. Rioux, D. P. B. McGovern, J. H. Cho, M. Georges, M. J. Daly, and J. C. Barrett. Fine-mapping inflammatory bowel disease loci to single-variant resolution.* Nature*, 547(7662):173–178, 07 2017. PMCID: PMC5511510.*

[37] *A. Mahajan, J. Wessel, S. M. Willems, W. Zhao, N. R. Robertson, A. Y. Chu, W. Gan, H. Kitajima, D. Taliun, N. W. Rayner, X. Guo, Y. Lu, M. Li, R. A. Jensen, Y. Hu, S. Huo, K. K. Lohman, W. Zhang, J. P. Cook, B. P. Prins, J. Flannick, N. Grarup, V. V. Trubetskoy, J. Kravic, Y. J. Kim, D. V. Rybin, H. Yaghootkar, M. Müller-Nurasyid, K. Meidtner, R. Li-Gao, T. V. Varga, J. Marten, J. Li, A. V. Smith, P. An, S. Ligthart, S. Gustafsson, G. Malerba, A. Demirkan, J. F. Tajes, V. Steinthorsdottir, M. Wuttke, C. Lecoeur, M. Preuss, L. F. Bielak, M. Graff, H. M. Highland, A. E. Justice, D. J. Liu, E. Marouli, G. M. Peloso, H. R. Warren, ExomeBP Consortium, MAGIC Consortium, GIANT Consortium, S. Afaq, S. Afzal, E. Ahlqvist, P. Almgren, N. Amin, L. B.*

Bang, A. G. Bertoni, C. Bombieri, J. Bork-Jensen, I. Brandslund, J. A. Brody, N. P. Burtt, M. Canouil, Y.-D. I. Chen, Y. S. Cho, C. Christensen, S. V. Eastwood, K.-U. Eckardt, K. Fischer, G. Gambaro, V. Giedraitis, M. L. Grove, H. G. de Haan, S. Hackinger, Y. Hai, S. Han, A. Tybjærg-Hansen, M.-F. Hivert, B. Isomaa, S. Jäger, M. E. Jørgensen, T. Jørgensen, A. Käräjämäki, B.-J. Kim, S. S. Kim, H. A. Koistinen, P. Kovacs, J. Kriebel, F. Kronenberg, K. Läll, L. A. Lange, J.-J. Lee, B. Lehne, H. Li, K.-H. Lin, A. Linneberg, C.-T. Liu, J. Liu, M. Loh, R. Mägi, V. Mamakou, R. McKean-Cowdin, G. Nadkarni, M. Neville, S. F. Nielsen, I. Ntalla, P. A. Peyser, W. Rathmann, K. Rice, S. S. Rich, L. Rode, O. Rolandsson, S. Schönherr, E. Selvin, K. S. Small, A. Stančáková, P. Surendran, K. D. Taylor, T. M. Teslovich, B. Thorand, G. Thorleifsson, A. Tin, A. Tönjes, A. Varbo, D. R. Witte, A. R. Wood, P. Yajnik, J. Yao, L. Yengo, R. Young, P. Amouyel, H. Boeing, E. Boerwinkle, E. P. Bottinger, R. Chowdhury, F. S. Collins, G. Dedoussis, A. Dehghan, P. Deloukas, M. M. Ferrario, J. Ferrières, J. C. Florez, P. Frossard, V. Gudnason, T. B. Harris, S. R. Heckbert, J. M. M. Howson, M. Ingelsson, S. Kathiresan, F. Kee, J. Kuusisto, C. Langenberg, L. J. Launer, C. M. Lindgren, S. Männistö, T. Meitinger, O. Melander, K. L. Mohlke, M. Moitry, A. D. Morris, A. D. Murray, R. de Mutsert, M. Orho-Melander, K. R. Owen, M. Perola, A. Peters, M. A. Province, A. Rasheed, P. M. Ridker, F. Rivadineira, F. R. Rosendaal, A. H. Rosengren, V. Salomaa, W. H.-H. Sheu, R. Sladek, B. H. Smith, K. Strauch, A. G. Uitterlinden, R. Varma, C. J. Willer, M. Blüher, A. S. Butterworth, J. C. Chambers, D. I. Chasman, J. Danesh, C. van Duijn, J. Dupuis, O. H. Franco, P. W. Franks, P. Froguel, H. Grallert, L. Groop, B.-G. Han, T. Hansen, A. T. Hattersley, C. Hayward, E. Ingelsson, S. L. R. Kardia, F. Karpe, J. S. Kooner, A. Köttgen, K. Kuulasmaa, M. Laakso, X. Lin, L. Lind, Y. Liu, R. J. F. Loos, J. Marchini, A. Metspalu, D. Mook-Kanamori, B. G. Nordestgaard, C. N. A. Palmer, J. S. Pankow, O. Pedersen, B. M. Psaty, R. Rauramaa, N. Sattar, M. B. Schulze, N. Soranzo, T. D. Spector, K. Stefansson, M. Stumvoll, U. Thorsteinsdottir, T. Tuomi, J. Tuomilehto, N. J. Wareham, J. G. Wilson, E. Zeggini, R. A. Scott, I. Barroso, T. M. Frayling, M. O. Goodarzi, J. B. Meigs, M. Boehnke, D. Saleheen, A. P. Morris, J. I. Rotter, and M. I. McCarthy. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet*, 50(4):559–571, 04 2018. PMCID: PMC5898373.*

[38] C. Benner, C. C. A. Spencer, A. S. Havulinna, V. Salomaa, S. Ripatti, and M. Pirinen. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics*, 32(10):1493–501, 05 2016. PMCID: PMC4866522.*

[39] B. Devlin and N. Risch. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics*, 29(2):311–22, Sep 1995. PMID: 8666377.*

[40] G. Wang, A. Sarkar, P. Carbonetto, and M. Stephens. A simple new approach to variable selection in regression, with application to genetic fine mapping. J.R. Statist. Soc. B*, 82:1273–1300, 2020.*

[41] K. K.-H. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shoresh, H. Whitton, R. J. H. Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos, P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature*, 518(7539):337–43, Feb 2015. PMCID: PMC4336207.*

[42] G. Kichaev, W.-Y. Yang, S. Lindstrom, F. Hormozdiari, E. Eskin, A. L. Price, P. Kraft, and B. Pasaniuc. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet*, 10(10):e1004722, Oct 2014. PMCID: PMC4214605.*

[43] W. Chen, S. K. McDonnell, S. N. Thibodeau, L. S. Tillmans, and D. J. Schaid. Incorporating Functional Annotations for Fine-Mapping Causal Variants in a Bayesian Framework Using Summary Statistics. Genetics*, 204(3):933–958, 11 2016. PMCID: PMC5105870.*

[44] H.-J. Westra, M. Martínez-Bonet, S. Onengut-Gumuscu, A. Lee, Y. Luo, N. Teslovich, J. Worthington, J. Martin, T. Huizinga, L. Klareskog, S. Rantapaa-Dahlqvist, W.-M. Chen, A. Quinlan, J. A. Todd, S. Eyre, P. A. Nigrovic, P. K. Gregersen, S. S. Rich, and S. Raychaudhuri. Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes. Nat Genet*, 50(10):1366–1374, 10 2018. PMCID: PMC6364548.*

[45] J. C. Ulirsch, C. A. Lareau, E. L. Bao, L. S. Ludwig, M. H. Guo, C. Benner, A. T. Satpathy, V. K. Kartha, R. M. Salem, J. N. Hirschhorn, H. K. Finucane, M. J. Aryee, J. D. Buenrostro, and V. G. Sankaran. Interrogation of human hematopoiesis at single-cell and single-variant resolution. Nat Genet*, 51(4):683–693, 04 2019. PMCID: PMC6441389.*

[46] O. Weissbrod, F. Hormozdiari, C. Benner, R. Cui, J. Ulirsch, S. Gazal, A. P. Schoech, B. van de Geijn, Y. Reshef, C. Márquez-Luna, L. O'Connor, M. Pirinen, H. K. Finucane, and A. L. Price. Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nat Genet, 52(12):1355–1363, 12 2020. PMCID: PMC7710571.

[47] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414):57–74, Sep 2012. PMCID: PMC3439153.

[48] G. Trynka, C. Sandor, B. Han, H. Xu, B. E. Stranger, X. S. Liu, and S. Raychaudhuri. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet, 45(2):124–30, Feb 2013. PMCID: PMC3826950.

[49] Roadmap Epigenomics Consortium, 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. M. 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, and M. Kellis. Integrative analysis of 111 reference human epigenomes. Nature, 518(7539):317–30, Feb 2015. PMCID: PMC4530010.

[50] P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, and M. Kircher. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res, 47(D1):D886–D894, 01 2019. PMCID: PMC6323892.

[51] D. Wang, S. Liu, J. Warrell, H. Won, X. Shi, F. C. P. Navarro, D. Clarke, M. Gu, P. Emani, Y. T. Yang, M. Xu, M. J. Gandal, S. Lou, J. Zhang, J. J. Park, C. Yan, S. K. Rhie, K. Manakongtreecheep, H. Zhou, A. Nathan, M. Peters, E. Mattei, D. Fitzgerald, T. Brunetti, J. Moore, Y. Jiang, K. Girdhar, G. E. Hoffman, S. Kalayci, Z. H. Gümüş, G. E. Crawford, PsychENCODE Consortium, P. Roussos, S. Akbarian, A. E. Jaffe, K. P. White, Z. Weng, N. Sestan, D. H. Geschwind, J. A. Knowles, and M. B. Gerstein. Comprehensive functional genomic resource and integrative model for the human brain. Science, 362(6420), 12 2018. PMCID: PMC6413328.

[52] PsychENCODE Consortium. Revealing the brain's molecular architecture. Science, 362(6420):1262–1263, Dec 2018. PMID: 30545881.

[53] M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel, M. Varghese, Y. Wang, A. W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran Losada, Z. Khan, J. Mleczko, Y. Xia, R. Dai, D. Wang, Y. T. Yang, M. Xu, K. Fish, P. R. Hof, J. Warrell, D. Fitzgerald, K. White, A. E. Jaffe, PsychENCODE Consortium, M. A. Peters, M. Gerstein, C. Liu, L. M. Iakoucheva, D. Pinto, and D. H. Geschwind. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science, 362(6420), 12 2018. PMCID: PMC6443102.

[54] GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science, 369(6509):1318–1330, 09 2020. PMCID: PMC7737656.

[55] S. Kim-Hellmuth, F. Aguet, M. Oliva, M. Muñoz-Aguirre, S. Kasela, V. Wucher, S. E. Castel, A. R. Hamel, A. Viñuela, A. L. Roberts, S. Mangul, X. Wen, G. Wang, A. N. Barbeira, D. Garrido-Martín, B. B. Nadel, Y. Zou, R. Bonazzola, J. Quan, A. Brown, A. Martinez-Perez, J. M. Soria, GTEx Consortium, G. Getz, E. T. Dermitzakis, K. S. Small, M. Stephens, H. S. Xi, H. K. Im, R. Guigó, A. V. Segrè, B. E. Stranger, K. G. Ardlie, and T. Lappalainen. Cell type-specific genetic regulation of gene expression across human tissues. Science, 369(6509), 09 2020. PMCID: PMC8051643.

[56] S. E. Castel, F. Aguet, P. Mohammadi, GTEx Consortium, K. G. Ardlie, and T. Lappalainen. A vast resource of allelic expression data spanning human tissues. Genome Biol, 21(1):234, 09 2020. PMCID: PMC7488534.

[57] *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):1190–5, Sep 2012. PMCID: PMC3771521.*

[58] *J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits.* Am J Hum Genet*, 94(4):559–73, Apr 2014. PMCID: PMC3980523.*

[59] *H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, RACI Consortium, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, and A. L. Price. Partitioning heritability by functional annotation using genome-wide association summary statistics.* Nat Genet*, 47(11):1228–35, Nov 2015. PMCID: PMC4626285.*

[60] *A. F. Pardiñas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D. Cameron, M. L. Hamshere, J. Han, L. Hubbard, A. Lynham, K. Mantripragada, E. Rees, J. H. MacCabe, S. A. McCarroll, B. T. Baune, G. Breen, E. M. Byrne, U. Dannlowski, T. C. Eley, C. Hayward, N. G. Martin, A. M. McIntosh, R. Plomin, D. J. Porteous, N. R. Wray, A. Caballero, D. H. Geschwind, L. M. Huckins, D. M. Ruderfer, E. Santiago, P. Sklar, E. A. Stahl, H. Won, E. Agerbo, T. D. Als, O. A. Andreassen, M. Bækvad-Hansen, P. B. Mortensen, C. B. Pedersen, A. D. Børglum, J. Bybjerg-Grauholm, S. Djurovic, N. Durmishi, M. G. Pedersen, V. Golimbet, J. Grove, D. M. Hougaard, M. Mattheisen, E. Molden, O. Mors, M. Nordentoft, M. Pejovic-Milovancevic, E. Sigurdsson, T. Silagadze, C. S. Hansen, K. Stefansson, H. Stefansson, S. Steinberg, S. Tosato, T. Werge, GERAD1 Consortium, CRESTAR Consortium, D. A. Collier, D. Rujescu, G. Kirov, M. J. Owen, M. C. O'Donovan, and J. T. R. Walters. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection.* Nat Genet*, 50(3):381–389, 03 2018. PMCID: PMC5918692.*

[61] *M. L. A. Hujoel, S. Gazal, F. Hormozdiari, B. van de Geijn, and A. L. Price. Disease Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient Sequence Age and Conserved Function across Species.* Am J Hum Genet*, 104(4):611–624, 04 2019. PMCID: PMC6451699.*

[62] *Y. Zou, P. Carbonetto, G. Wang, and M. Stephens. Fine-mapping from summary data with the "Sum of Single Effects" model.* bioRxiv*, 2021.*

[63] *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):667–80, May 2013. PMCID: PMC3644637.*

[64] *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. PMCID: PMC4022491.*

[65] *A. Dobbyn, L. M. Huckins, J. Boocock, L. G. Sloofman, B. S. Glicksberg, C. Giambartolomei, G. E. Hoffman, T. M. Perumal, K. Girdhar, Y. Jiang, T. Raj, D. M. Ruderfer, R. S. Kramer, D. Pinto, CommonMind Consortium, S. Akbarian, P. Roussos, E. Domenici, B. Devlin, P. Sklar, E. A. Stahl, and S. K. Sieberts. Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS.* Am J Hum Genet*, 102(6):1169–1184, 06 2018. PMCID: PMC5993513.*

[66] *C. Giambartolomei, J. Zhenli Liu, W. Zhang, M. Hauberg, H. Shi, J. Boocock, J. Pickrell, A. E. Jaffe, CommonMind Consortium, B. Pasaniuc, and P. Roussos. A Bayesian framework for multiple trait colocalization from summary association statistics.* Bioinformatics*, 34(15):2538–2545, 08 2018. PMCID: PMC6061859.*

[67] *C. N. Foley, J. R. Staley, P. G. Breen, B. B. Sun, P. D. W. Kirk, S. Burgess, and J. M. M. Howson. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits.* Nat Commun*, 12(1):764, 02 2021. PMCID: PMC7858636.*

[68] *F. Hormozdiari, M. van de Bunt, A. V. Segrè, X. Li, J. W. J. Joo, M. Bilow, J. H. Sul, S. Sankararaman, B. Pasaniuc, and E. Eskin. Colocalization of GWAS and eQTL Signals Detects Target Genes.* Am J Hum Genet*, 99(6):1245–1260, Dec 2016. PMCID: PMC5142122.*

[69] A. Zhu, N. Matoba, E. P. Wilson, A. L. Tapia, Y. Li, J. G. Ibrahim, J. L. Stein, and M. I. Love. MRLocus: Identifying causal genes mediating a trait through Bayesian estimation of allelic heterogeneity. PLoS Genet, 17(4):e1009455, 04 2021. PMCID: PMC8084342.

[70] 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):481–7, 05 2016. PMID: 27019110.

[71] A. Hukku, M. Pividori, F. Luca, R. Pique-Regi, H. K. Im, and X. Wen. Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. Am J Hum Genet, 108(1):25–35, 01 2021. PMCID: PMC7820626.

[72] X. Wen, R. Pique-Regi, and F. Luca. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. PLoS Genet, 13(3):e1006646, Mar 2017. PMCID: PMC5363995.

[73] M. Pividori, P. S. Rajagopal, A. Barbeira, Y. Liang, O. Melia, L. Bastarache, Y. Park, G. Consortium, X. Wen, and H. K. Im. PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci Adv, 6(37), 09 2020. PMID: 32917697.

[74] M. Fromer, P. Roussos, S. K. Sieberts, J. S. Johnson, D. H. Kavanagh, T. M. Perumal, D. M. Ruderfer, E. C. Oh, A. Topol, H. R. Shah, L. L. Klei, R. Kramer, D. Pinto, Z. H. Gümüş, A. E. Cicek, K. K. Dang, A. Browne, C. Lu, L. Xie, B. Readhead, E. A. Stahl, J. Xiao, M. Parvizi, T. Hamamsy, J. F. Fullard, Y.-C. Wang, M. C. Mahajan, J. M. J. Derry, J. T. Dudley, S. E. Hemby, B. A. Logsdon, K. Talbot, T. Raj, D. A. Bennett, P. L. De Jager, J. Zhu, B. Zhang, P. F. Sullivan, A. Chess, S. M. Purcell, L. A. Shinobu, L. M. Mangravite, H. Toyoshiba, R. E. Gur, C.-G. Hahn, D. A. Lewis, V. Haroutunian, M. A. Peters, B. K. Lipska, J. D. Buxbaum, E. E. Schadt, K. Hirai, K. Roeder, K. J. Brennand, N. Katsanis, E. Domenici, B. Devlin, and P. Sklar. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat Neurosci, 19(11):1442–1453, 11 2016. PMCID: PMC5083142.

[75] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, GTEx Consortium, 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):1091–8, Sep 2015. PMCID: PMC4552594.

[76] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. J. H. Penninx, R. Jansen, E. J. C. de Geus, D. I. Boomsma, F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. Lehtimäki, E. Raitoharju, M. Kähönen, I. Seppälä, 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):245–52, Mar 2016. PMCID: PMC4767558.

[77] A. N. Barbeira, M. Pividori, J. Zheng, H. E. Wheeler, D. L. Nicolae, and H. K. Im. Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet, 15(1):e1007889, 01 2019. PMCID: PMC6358100.

[78] H. Feng, N. Mancuso, A. Gusev, A. Majumdar, M. Major, B. Pasaniuc, and P. Kraft. Leveraging expression from multiple tissues using sparse canonical correlation analysis and aggregate tests improves the power of transcriptome-wide association studies. PLoS Genet, 17(4):e1008973, 04 2021. PMCID: PMC8057593.

[79] M. Wainberg, N. Sinnott-Armstrong, N. Mancuso, A. N. Barbeira, D. A. Knowles, D. Golan, R. Ermel, A. Ruusalepp, T. Quertermous, K. Hao, J. L. M. Björkegren, H. K. Im, B. Pasaniuc, M. A. Rivas, and A. Kundaje. Opportunities and challenges for transcriptome-wide association studies. Nat Genet, 51(4):592–599, 04 2019. PMCID: PMC6777347.

[80] B. Li, Y. Veturi, A. Verma, Y. Bradford, E. S. Daar, R. M. Gulick, S. A. Riddler, G. K. Robbins, J. L. Lennox, D. W. Haas, et al. Tissue specificity-aware TWAS (TSA-TWAS) framework identifies novel associations with metabolic, immunologic, and virologic traits in HIV-positive adults. PLoS genetics, 17(4):e1009464, 2021.

[81] A. Hukku, M. G. Sampson, F. Luca, R. Pique-Regi, and X. Wen. Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. Am J Hum Genet, 109(5):825–837, May 2022. PMID: 35523146.

[82] International Schizophrenia Consortium, S. M. Purcell, N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan, and P. Sklar. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature, 460(7256):748–52, Aug 2009. PMCID: PMC3912837.

[83] F. Privé, B. J. Vilhjálmsson, H. Aschard, and M. G. B. Blum. Making the Most of Clumping and Thresholding for Polygenic Scores. Am J Hum Genet, 105(6):1213–1221, 12 2019. PMCID: PMC6904799.

[84] F. Privé, J. Arbel, and B. J. Vilhjálmsson. LDpred2: better, faster, stronger. Bioinformatics, Dec 2020. PMCID: PMC8016455.

[85] C. Márquez-Luna, S. Gazal, P.-R. Loh, S. S. Kim, N. Furlotte, A. Auton, 23andMe Research Team, and A. L. Price. Incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. Nat Commun, 12(1):6052, 10 2021. PMCID: PMC8523709.

[86] T. S. H. Mak, R. M. Porsch, S. W. Choi, X. Zhou, and P. C. Sham. Polygenic scores via penalized regression on summary statistics. Genet Epidemiol, 41(6):469–480, 09 2017. PMID: 28480976.

[87] L. R. Lloyd-Jones, J. Zeng, J. Sidorenko, L. Yengo, G. Moser, K. E. Kemper, H. Wang, Z. Zheng, R. Magi, T. Esko, A. Metspalu, N. R. Wray, M. E. Goddard, J. Yang, and P. M. Visscher. Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat Commun, 10(1):5086, 11 2019. PMCID: PMC6841727.

[88] Q. Zhang, F. Privé, B. Vilhjálmsson, and D. Speed. Improved genetic prediction of complex traits from individual-level data or summary statistics. Nat Commun, 12(1):4192, 07 2021. PMCID: PMC8263809.

[89] J. Pattee and W. Pan. Penalized regression and model selection methods for polygenic scores on summary statistics. PLoS Comput Biol, 16(10):e1008271, 10 2020. PMCID: PMC7553329.

[90] G. Ni, J. Zeng, J. A. Revez, Y. Wang, Z. Zheng, T. Ge, R. Restuadi, J. Kiewa, D. R. Nyholt, J. R. I. Coleman, J. W. Smoller, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, J. Yang, P. M. Visscher, and N. R. Wray. A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. Biol Psychiatry, 90(9):611–620, 11 2021. PMCID: PMC8500913.

[91] D. Curtis. Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. Psychiatr Genet, 28(5):85–89, 10 2018. PMID: 30160659.

[92] S. Yang and X. Zhou. PGS-server: accuracy, robustness and transferability of polygenic score methods for biobank scale studies. Brief Bioinform, 23(2), 03 2022. PMID: 35193147.

[93] Y. Hu, Q. Lu, W. Liu, Y. Zhang, M. Li, and H. Zhao. Joint modeling of genetically correlated diseases and functional annotations increases accuracy of polygenic risk prediction. PLoS Genet, 13(6):e1006836, Jun 2017. PMCID: PMC5482506.

[94] T. Amariuta, K. Ishigaki, H. Sugishita, T. Ohta, M. Koido, K. K. Dey, K. Matsuda, Y. Murakami, A. L. Price, E. Kawakami, C. Terao, and S. Raychaudhuri. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nat Genet, 52(12):1346–1354, 12 2020. PMCID: PMC8049522.

[95] C. Márquez-Luna, P.-R. Loh, South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium, and A. L. Price. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol, 41(8):811–823, 12 2017. PMCID: PMC5726434.

[96] D. J. Lawson, G. Hellenthal, S. Myers, and D. Falush. Inference of population structure using dense haplotype data. PLoS Genet, 8(1):e1002453, Jan 2012. PMCID: PMC3266881.

[97] Y. Guan. Detecting structure of haplotypes and local ancestry. Genetics, 196(3):625–42, Mar 2014. PMCID: PMC3948796.

[98] D. Marnetto, K. Pärna, K. Läll, L. Molinaro, F. Montinaro, T. Haller, M. Metspalu, R. Mägi, K. Fischer, and L. Pagani. Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. Nat Commun, 11(1):1628, 04 2020. PMCID: PMC7118071.

[99] *A. R. Martin, M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, and M. J. Daly. Clinical use of current polygenic risk scores may exacerbate health disparities.* Nat Genet*, 51(4):584–591, 04 2019. PMCID: PMC6563838.*

[100] *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. S. Celestino-Soper, M. Choi, E. L. Crawford, L. Davis, N. R. D. 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. Günel, 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, Jr, 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):863–85, Jun 2011. PMCID: PMC3939065.*

[101] *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. Günel, 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):237–41, Apr 2012. PMCID: PMC3667984.*

[102] *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, Jr, 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):242–5, Apr 2012. PMCID: PMC3613847.*

[103] *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.*

[104] *S. De Rubeis, X. He, A. P. Goldberg, C. S. Poultney, K. Samocha, A. E. Cicek, Y. Kou, L. Liu, M. Fromer, S. Walker, T. Singh, L. Klei, J. Kosmicki, F. Shih-Chen, B. Aleksic, M. Biscaldi, P. F. Bolton, J. M. Brownfeld, J. Cai, N. G. Campbell, A. Carracedo, M. H. Chahrour, A. G. Chiocchetti, H. Coon, E. L. Crawford, S. R. Curran, G. Dawson, E. Duketis, B. A. Fernandez, L. Gallagher, E. Geller, S. J. Guter, R. S. Hill, J. Ionita-Laza, P. Jimenz Gonzalez, H. Kilpinen, S. M. Klauck, A. Kolevzon, I. Lee, I. Lei, J. Lei, T. Lehtimäki, C.-F. Lin, A. Ma'ayan, C. R. Marshall, A. L. McInnes, B. Neale, M. J. Owen, N. Ozaki, M. Parellada, J. R. Parr, S. Purcell, K. Puura, D. Rajagopalan, K. Rehnström, A. Reichenberg, A. Sabo, M. Sachse, S. J. Sanders, C. Schafer, M. Schulte-Rüther, D. Skuse, C. Stevens, P. Szatmari, K. Tammimies, O. Valladares, A. Voran, W. Li-San, L. A. Weiss, A. J. Willsey, T. W. Yu, R. K. C. Yuen, DDD Study, Homozygosity Mapping Collaborative for Autism, UK10K Consortium, E. H. Cook, C. M. Freitag, M. Gill, C. M. Hultman, T. Lehner, A. Palotie, G. D. Schellenberg, P. Sklar, M. W. State, J. S. Sutcliffe, C. A. Walsh, S. W. Scherer, M. E. Zwick, J. C. Barett, D. J. Cutler, K. Roeder, B. Devlin, M. J. Daly, and J. D. Buxbaum. Synaptic, transcriptional and chromatin genes disrupted in autism.* Nature*, 515(7526):209–15, Nov 2014. PMCID: PMC4402723.*

[105] *A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan, and M. W. State. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism.* Cell*, 155(5):997–1007, Nov 2013. PMCID: PMC3995413.*

[106] *J. Cotney, R. A. Muhle, S. J. Sanders, L. Liu, A. J. Willsey, W. Niu, W. Liu, L. Klei, J. Lei, J. Yin, S. K. Reilly, A. T. Tebbenkamp, C. Bichsel, M. Pletikos, N. Sestan, K. Roeder, M. W. State, B. Devlin, and J. P.*

Noonan. The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat Commun*, 6:6404, Mar 2015. PMCID: PMC4355952.*

[107] J. D. Weissenkampen, Y. Jiang, S. Eckert, B. Jiang, B. Li, and D. J. Liu. Methods for the Analysis and Interpretation for Rare Variants Associated with Complex Traits. Curr Protoc Hum Genet*, 101(1):e83, 04 2019. PMCID: PMC6455968.*

[108] T. Lappalainen, A. J. Scott, M. Brandt, and I. M. Hall. Genomic Analysis in the Age of Human Genome Sequencing. Cell*, 177(1):70–84, 03 2019. PMCID: PMC6532068.*

[109] O. Bocher and E. Génin. Rare variant association testing in the non-coding genome. Hum Genet*, 139(11):1345–1362, Nov 2020. PMID: 32500240.*

[110] D. M. Werling, H. Brand, J.-Y. An, M. R. Stone, L. Zhu, J. T. Glessner, R. L. Collins, S. Dong, R. M. Layer, E. Markenscoff-Papadimitriou, A. Farrell, G. B. Schwartz, H. Z. Wang, B. B. Currall, X. Zhao, J. Dea, C. Duhn, C. A. Erdman, M. C. Gilson, R. Yadav, R. E. Handsaker, S. Kashin, L. Klei, J. D. Mandell, T. J. Nowakowski, Y. Liu, S. Pochareddy, L. Smith, M. F. Walker, M. J. Waterman, X. He, A. R. Kriegstein, J. L. Rubenstein, N. Sestan, S. A. McCarroll, B. M. Neale, H. Coon, A. J. Willsey, J. D. Buxbaum, M. J. Daly, M. W. State, A. R. Quinlan, G. T. Marth, K. Roeder, B. Devlin, M. E. Talkowski, and S. J. Sanders. An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. Nat Genet*, 50(5):727–736, 04 2018. PMCID: PMC5961723.*

[111] L. Zhu, J. Lei, B. Devlin, and K. Roeder. TESTING HIGH-DIMENSIONAL COVARIANCE MATRICES, WITH APPLICATION TO DETECTING SCHIZOPHRENIA RISK GENES. Ann Appl Stat*, 11(3):1810–1831, Sep 2017. PMCID: PMC5655846.*

[112] C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, and J. Marchini. The UK Biobank resource with deep phenotyping and genomic data. Nature*, 562(7726):203–209, 10 2018. PMCID: PMC6786975.*

[113] J. Lei. Cross-validation with confidence. Journal of the American Statistical Association*, 115(532):1978–1997, 2020.*

[114] N. Kissel and L. Mentch. Forward Stability and Model Path Selection. arXiv preprint arXiv:2103.03462*, 2021.*

[115] N. LaPierre, K. Taraszka, H. Huang, R. He, F. Hormozdiari, and E. Eskin. Identifying causal variants by fine mapping across multiple studies. PLoS Genet*, 17(9):e1009733, 09 2021. PMCID: PMC8491908.*

[116] U. Ohler, G.-c. Liao, H. Niemann, and G. M. Rubin. Computational analysis of core promoters in the Drosophila genome. Genome Biol*, 3(12):RESEARCH0087, 2002. PMCID: PMC151189.*

[117] L. Vo Ngoc, C. Y. Huang, C. J. Cassidy, C. Medrano, and J. T. Kadonaga. Identification of the human DPR core promoter element using machine learning. Nature*, 585(7825):459–463, 09 2020. PMCID: PMC7501168.*

[118] M. Kellis, B. Wold, M. P. Snyder, B. E. Bernstein, A. Kundaje, G. K. Marinov, L. D. Ward, E. Birney, G. E. Crawford, J. Dekker, I. Dunham, L. L. Elnitski, P. J. Farnham, E. A. Feingold, M. Gerstein, M. C. Giddings, D. M. Gilbert, T. R. Gingeras, E. D. Green, R. Guigo, T. Hubbard, J. Kent, J. D. Lieb, R. M. Myers, M. J. Pazin, B. Ren, J. A. Stamatoyannopoulos, Z. Weng, K. P. White, and R. C. Hardison. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A*, 111(17):6131–8, Apr 2014. PMCID: PMC4035993.*

[119] S. Gazal, H. K. Finucane, N. A. Furlotte, P.-R. Loh, P. F. Palamara, X. Liu, A. Schoech, B. Bulik-Sullivan, B. M. Neale, A. Gusev, and A. L. Price. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat Genet*, 49(10):1421–1427, Oct 2017. PMCID: PMC6133304.*

[120] C. Quick, X. Wen, G. Abecasis, M. Boehnke, and H. M. Kang. Integrating comprehensive functional annotations to boost power and accuracy in gene-based association analysis. PLoS Genet*, 16(12):e1009060, 12 2020. PMCID: PMC7737906.*

[121] A. Amlie-Wolf, M. Tang, E. E. Mlynarski, P. P. Kuksa, O. Valladares, Z. Katanic, D. Tsuang, C. D. Brown, G. D. Schellenberg, and L.-S. Wang. INFERNO: inferring the molecular mechanisms of noncoding genetic variants. Nucleic Acids Res, 46(17):8740–8753, 09 2018. PMCID: PMC6158604.

[122] A. Amlie-Wolf, P. P. Kuksa, C.-Y. Lee, E. Mlynarski, Y. Y. Leung, and L.-S. Wang. Using INFERNO to Infer the Molecular Mechanisms Underlying Noncoding Genetic Associations. Methods Mol Biol, 2254:73–91, 2021. PMID: 33326071.

[123] X. Li, Z. Li, H. Zhou, S. M. Gaynor, Y. Liu, H. Chen, R. Sun, R. Dey, D. K. Arnett, S. Aslibekyan, C. M. Ballantyne, L. F. Bielak, J. Blangero, E. Boerwinkle, D. W. Bowden, J. G. Broome, M. P. Conomos, A. Correa, L. A. Cupples, J. E. Curran, B. I. Freedman, X. Guo, G. Hindy, M. R. Irvin, S. L. R. Kardia, S. Kathiresan, A. T. Khan, C. L. Kooperberg, C. C. Laurie, X. S. Liu, M. C. Mahaney, A. W. Manichaikul, L. W. Martin, R. A. Mathias, S. T. McGarvey, B. D. Mitchell, M. E. Montasser, J. E. Moore, A. C. Morrison, J. R. O'Connell, N. D. Palmer, A. Pampana, J. M. Peralta, P. A. Peyser, B. M. Psaty, S. Redline, K. M. Rice, S. S. Rich, J. A. Smith, H. K. Tiwari, M. Y. Tsai, R. S. Vasan, F. F. Wang, D. E. Weeks, Z. Weng, J. G. Wilson, L. R. Yanek, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Lipids Working Group, B. M. Neale, S. R. Sunyaev, G. R. Abecasis, J. I. Rotter, C. J. Willer, G. M. Peloso, P. Natarajan, and X. Lin. Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. Nat Genet, 52(9):969–983, 09 2020. PMCID: PMC7483769.

[124] S. Gazal, O. Weissbrod, F. Hormozdiari, K. K. Dey, J. Nasser, K. A. Jagadeesh, D. J. Weiner, H. Shi, C. P. Fulco, L. J. O'Connor, B. Pasaniuc, J. M. Engreitz, and A. L. Price. Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. Nat Genet, Jun 2022. PMID: 35668300.

[125] M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur, and Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536(7616):285–91, 08 2016. PMCID: PMC5018207.

[126] K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski, Genome Aggregation Database Consortium, B. M. Neale, M. J. Daly, and D. G. MacArthur. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature, 581(7809):434–443, 05 2020. PMCID: PMC7334197.

[127] J. M. Havrilla, B. S. Pedersen, R. M. Layer, and A. R. Quinlan. A map of constrained coding regions in the human genome. Nat Genet, 51(1):88–95, 01 2019. PMCID: PMC6589356.

[128] S. Chen, R. Fragoza, L. Klei, Y. Liu, J. Wang, K. Roeder, B. Devlin, and H. Yu. An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. Nat Genet, 50(7):1032–1040, 07 2018. PMCID: PMC6314957.

[129] Y. Liu, Y. Liang, A. E. Cicek, Z. Li, J. Li, R. A. Muhle, M. Krenzer, Y. Mei, Y. Wang, N. Knoblauch, J. Morrison, S. Zhao, Y. Jiang, E. Geller, I. Ionita-Laza, J. Wu, K. Xia, J. P. Noonan, Z. S. Sun, and X. He. A Statistical Framework for Mapping Risk Genes from De Novo Mutations in Whole-Genome-Sequencing Studies. Am J Hum Genet, 102(6):1031–1047, 06 2018. PMCID: PMC5992125.

[130] C. Villani. Optimal transport: old and new, volume 338. Springer, 2009.

[131] G. Peyré, M. Cuturi, et al. Computational optimal transport: With applications to data science. Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.*

[132] V. M. Panaretos and Y. Zemel. An invitation to statistics in Wasserstein space*. Springer Nature, 2020.*

[133] J. Bryois, D. Calini, W. Macnair, L. Foo, E. Urich, W. Ortmann, V. A. Iglesias, S. Selvaraj, E. Nutma, M. Marzin, S. Amor, A. Williams, G. Castelo-Branco, V. Menon, P. De Jager, and D. Malhotra. Cell-type specific cis-eQTLs in eight brain cell-types identifies novel risk genes for human brain disorders. medRxiv*, 2021.*

[134] E. M. Weeks, J. C. Ulirsch, N. Y. Cheng, B. L. Trippe, R. S. Fine, J. Miao, T. A. Patwardhan, M. Kanai, J. Nasser, C. P. Fulco, K. C. Tashman, F. Aguet, T. Li, J. Ordovas-Montanes, C. S. Smillie, M. Biton, A. K. Shalek, A. N. Ananthakrishnan, R. J. Xavier, A. Regev, R. M. Gupta, K. Lage, K. G. Ardlie, J. N. Hirschhorn, E. S. Lander, J. M. Engreitz, and H. K. Finucane. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. medRxiv*, 2020*.

[135] 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, Mar 2014. PMCID: PMC4016412.*

[136] L. Liu, J. Lei, and K. Roeder. NETWORK ASSISTED ANALYSIS TO REVEAL THE GENETIC BASIS OF AUTISM. Ann Appl Stat*, 9(3):1571–1600, 2015. PMCID: PMC4851445.*

[137] X. Li, G. Yung, H. Zhou, R. Sun, Z. Li, K. Hou, M. J. Zhang, Y. Liu, T. Arapoglou, C. Wang, I. Ionita-Laza, and X. Lin. A multi-dimensional integrative scoring framework for predicting functional variants in the human genome. Am J Hum Genet*, 109(3):446–456, 03 2022. PMCID: PMC8948160.*

[138] D. W. Yao, L. J. O'Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat Genet*, 52(6):626–633, 06 2020. PMCID: PMC7276299.*

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

[140] A. T. Ghanbarian and L. D. Hurst. Neighboring Genes Show Correlated Evolution in Gene Expression. Mol Biol Evol*, 32(7):1748–66, Jul 2015. PMCID: PMC4476153.*

[141] O. Symmons, V. V. Uslu, T. Tsujimura, S. Ruf, S. Nassari, W. Schwarzer, L. Ettwiller, and F. Spitz. Functional and topological characteristics of mammalian regulatory domains. Genome Res*, 24(3):390–400, Mar 2014. PMCID: PMC3941104.*

[142] M. Ebisuya, T. Yamamoto, M. Nakajima, and E. Nishida. Ripples from neighbouring transcription. Nat Cell Biol*, 10(9):1106–13, Sep 2008. PMID: 19160492.*

[143] C. Wadsworth, F. Vera, and C. Piech. Achieving fairness through adversarial learning: an application to recidivism prediction. arXiv preprint arXiv:1807.00199*, 2018.*

[144] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork. Learning fair representations. In International conference on machine learning*, pages 325–333. PMLR, 2013.*

[145] D. Madras, E. Creager, T. Pitassi, and R. Zemel. Learning adversarially fair and transferable representations. In International Conference on Machine Learning*, pages 3384–3393. PMLR, 2018.*

[146] M. Takada, T. Suzuki, and H. Fujisawa. Independently interpretable lasso: A new regularizer for sparse regression with uncorrelated variables. In International Conference on Artificial Intelligence and Statistics*, pages 454–463. PMLR, 2018.*

[147] Y. E. Bae, L. Wu, and C. Wu. InTACT: An adaptive and powerful framework for joint-tissue transcriptome-wide association studies. Genet Epidemiol*, 45(8):848–859, 12 2021. PMCID: PMC8604767.*

[148] *PsychENCODE Consortium, 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. P. 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. Y. 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):1707–12, Dec 2015. PMCID: PMC4675669.*

[149] *B. Efron. The estimation of prediction error: covariance penalties and cross-validation.* Journal of the American Statistical Association*, 99(467):619–632, 2004.*

[150] *J. Leiner, B. Duan, L. Wasserman, and A. Ramdas. Data blurring: sample splitting a single sample.* arXiv preprint arXiv:2112.11079*, 2021.*

[151] *D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A. G. Taliun, A. Corvelo, S. M. Gogarten, H. M. Kang, A. N. Pitsillides, J. LeFaive, S.-B. Lee, X. Tian, B. L. Browning, S. Das, A.-K. Emde, W. E. Clarke, D. P. Loesch, A. C. Shetty, T. W. Blackwell, A. V. Smith, Q. Wong, X. Liu, M. P. Conomos, D. M. Bobo, F. Aguet, C. Albert, A. Alonso, K. G. Ardlie, D. E. Arking, S. Aslibekyan, P. L. Auer, J. Barnard, R. G. Barr, L. Barwick, L. C. Becker, R. L. Beer, E. J. Benjamin, L. F. Bielak, J. Blangero, M. Boehnke, D. W. Bowden, J. A. Brody, E. G. Burchard, B. E. Cade, J. F. Casella, B. Chalazan, D. I. Chasman, Y.-D. I. Chen, M. H. Cho, S. H. Choi, M. K. Chung, C. B. Clish, A. Correa, J. E. Curran, B. Custer, D. Darbar, M. Daya, M. de Andrade, D. L. DeMeo, S. K. Dutcher, P. T. Ellinor, L. S. Emery, C. Eng, D. Fatkin, T. Fingerlin, L. Forer, M. Fornage, N. Franceschini, C. Fuchsberger, S. M. Fullerton, S. Germer, M. T. Gladwin, D. J. Gottlieb, X. Guo, M. E. Hall, J. He, N. L. Heard-Costa, S. R. Heckbert, M. R. Irvin, J. M. Johnsen, A. D. Johnson, R. Kaplan, S. L. R. Kardia, T. Kelly, S. Kelly, E. E. Kenny, D. P. Kiel, R. Klemmer, B. A. Konkle, C. Kooperberg, A. Köttgen, L. A. Lange, J. Lasky-Su, D. Levy, X. Lin, K.-H. Lin, C. Liu, R. J. F. Loos, L. Garman, R. Gerszten, S. A. Lubitz, K. L. Lunetta, A. C. Y. Mak, A. Manichaikul, A. K. Manning, R. A. Mathias, D. D. McManus, S. T. McGarvey, J. B. Meigs, D. A. Meyers, J. L. Mikulla, M. A. Minear, B. D. Mitchell, S. Mohanty, M. E. Montasser, C. Montgomery, A. C. Morrison, J. M. Murabito, A. Natale, P. Natarajan, S. C. Nelson, K. E. North, J. R. O'Connell, N. D. Palmer, N. Pankratz, G. M. Peloso, P. A. Peyser, J. Pleiness, W. S. Post, B. M. Psaty, D. C. Rao, S. Redline, A. P. Reiner, D. Roden, J. I. Rotter, I. Ruczinski, C. Sarnowski, S. Schoenherr, D. A. Schwartz, J.-S. Seo, S. Seshadri, V. A. Sheehan, W. H. Sheu, M. B. Shoemaker, N. L. Smith, J. A. Smith, N. Sotoodehnia, A. M. Stilp, W. Tang, K. D. Taylor, M. Telen, T. A. Thornton, R. P. Tracy, D. J. Van Den Berg, R. S. Vasan, K. A. Viaud-Martinez, S. Vrieze, D. E. Weeks, B. S. Weir, S. T. Weiss, L.-C. Weng, C. J. Willer, Y. Zhang, X. Zhao, D. K. Arnett, A. E. Ashley-Koch, K. C. Barnes, E. Boerwinkle, S. Gabriel, R. Gibbs, K. M. Rice, S. S. Rich, E. K. Silverman, P. Qasba, W. Gan, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, G. J. Papanicolaou, D. A. Nickerson, S. R. Browning, M. C. Zody, S. Zöllner, J. G. Wilson, L. A. Cupples, C. C. Laurie, C. E. Jaquish, R. D. Hernandez, T. D. O'Connor, and G. R. Abecasis. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.* Nature*, 590(7845):290–299, 02 2021. PMCID: PMC7875770.*

[152] *R. A. Power, S. Kyaga, R. Uher, J. H. MacCabe, N. Långström, M. Landen, P. McGuffin, C. M. Lewis, P. Lichtenstein, and A. C. Svensson. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings.* JAMA Psychiatry*, 70(1):22–30, Jan 2013. PMID: 23147713.*

[153] *A. Reichenberg, M. Cederlöf, A. McMillan, M. Trzaskowski, O. Kapra, E. Fruchter, K. Ginat, M. Davidson, M. Weiser, H. Larsson, R. Plomin, and P. Lichtenstein. Discontinuity in the genetic and environmental causes of the intellectual disability spectrum.* Proc Natl Acad Sci U S A*, 113(4):1098–103, 01 2016. PMCID: PMC4743770.*

[154] *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. PMCID: PMC4085641.*

[155] *P. L. Auer and G. Lettre. Rare variant association studies: considerations, challenges and opportunities.* Genome Med*, 7(1):16, 2015. PMCID: PMC4337325.*

[156] *L. Moutsianas, V. Agarwala, C. Fuchsberger, J. Flannick, M. A. Rivas, K. J. Gaulton, P. K. Albers, GoT2D Consortium, G. McVean, M. Boehnke, D. Altshuler, and M. I. McCarthy. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease.* PLoS Genet*, 11(4):e1005165, Apr 2015. PMCID: PMC4407972.*

[157] *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. PMCID: PMC3135811.*

[158] *S. Lee, M. C. Wu, and X. Lin. Optimal tests for rare variant effects in sequencing association studies.* Biostatistics*, 13(4):762–75, Sep 2012. PMCID: PMC3440237.*

[159] *W. Pan, J. Kim, Y. Zhang, X. Shen, and P. Wei. A powerful and adaptive association test for rare variants.* Genetics*, 197(4):1081–95, Aug 2014. PMCID: PMC4125385.*

[160] *T. Hasegawa, K. Kojima, Y. Kawai, K. Misawa, T. Mimori, and M. Nagasaki. AP-SKAT: highly-efficient genome-wide rare variant association test.* BMC Genomics*, 17(1):745, Sep 2016. PMCID: PMC5031335.*

[161] *R. Schweiger, O. Weissbrod, E. Rahmani, M. Müller-Nurasyid, S. Kunze, C. Gieger, M. Waldenberger, S. Rosset, and E. Halperin. RL-SKAT: An Exact and Efficient Score Test for Heritability and Set Tests.* Genetics*, 207(4):1275–1283, 12 2017. PMCID: PMC5714447.*

[162] *Z. He, Y. Le Guen, L. Liu, J. Lee, S. Ma, A. C. Yang, X. Liu, J. Rutledge, P. M. Losada, B. Song, M. E. Belloy, R. R. Butler, 3rd, F. M. Longo, H. Tang, E. C. Mormino, T. Wyss-Coray, M. D. Greicius, and I. Ionita-Laza. Genome-wide analysis of common and rare variants via multiple knockoffs at biobank scale, with an application to Alzheimer disease genetics.* Am J Hum Genet*, 108(12):2336–2353, 12 2021. PMCID: PMC8715147.*

[163] *Z. He, L. Liu, C. Wang, Y. Le Guen, J. Lee, S. Gogarten, F. Lu, S. Montgomery, H. Tang, E. K. Silverman, M. H. Cho, M. Greicius, and I. Ionita-Laza. Identification of putative causal loci in whole-genome sequencing data via knockoff statistics.* Nat Commun*, 12(1):3152, 05 2021. PMCID: PMC8149672.*

[164] *D. Xu, C. Wang, K. Kiryluk, J. D. Buxbaum, and I. Ionita-Laza. Co-localization between Sequence Constraint and Epigenomic Information Improves Interpretation of Whole-Genome Sequencing Data.* Am J Hum Genet*, 106(4):513–524, 04 2020. PMCID: PMC7118583.*

[165] *Z. He, B. Xu, J. Buxbaum, and I. Ionita-Laza. A genome-wide scan statistic framework for whole-genome sequence data analysis.* Nat Commun*, 10(1):3018, 07 2019. PMCID: PMC6616627.*

[166] *Y. Yang, Q. Sun, L. Huang, J. G. Broome, A. Correa, A. Reiner, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, L. M. Raffield, Y. Yang, and Y. Li. eSCAN: scan regulatory regions for aggregate association testing using whole-genome sequencing data.* Brief Bioinform*, 23(1), 01 2022. PMCID: PMC8898002.*

[167] *T. N. Turner, F. Hormozdiari, M. H. Duyzend, S. A. McClymont, P. W. Hook, I. Iossifov, A. Raja, C. Baker, K. Hoekzema, H. A. Stessman, M. C. Zody, B. J. Nelson, J. Huddleston, R. Sandstrom, J. D. Smith, D. Hanna, J. M. Swanson, E. M. Faustman, M. J. Bamshad, J. Stamatoyannopoulos, D. A. Nickerson, A. S. McCallion, R. Darnell, and E. E. Eichler. Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA.* Am J Hum Genet*, 98(1):58–74, Jan 2016. PMCID: PMC4716689.*

[168] *T. N. Turner, B. P. Coe, D. E. Dickel, K. Hoekzema, B. J. Nelson, M. C. Zody, Z. N. Kronenberg, F. Hormozdiari, A. Raja, L. A. Pennacchio, R. B. Darnell, and E. E. Eichler. Genomic Patterns of De Novo Mutation in Simplex Autism.* Cell*, 171(3):710–722.e12, Oct 2017. PMCID: PMC5679715.*

[169] *R. K. C. 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. C. 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):185–91, Feb 2015. PMID: 25621899.*

[170] *W. M. Brandler, D. Antaki, M. Gujral, M. L. Kleiber, J. Whitney, M. S. Maile, O. Hong, T. R. Chapman, S. Tan, P. Tandon, T. Pang, S. C. Tang, K. K. Vaux, Y. Yang, E. Harrington, S. Juul, D. J. Turner, B. Thiruvahindrapuram, G. Kaur, Z. Wang, S. F. Kingsmore, J. G. Gleeson, D. Bisson, B. Kakaradov, A. Telenti, J. C. Venter, R. Corominas, C. Toma, B. Cormand, I. Rueda, S. Guijarro, K. S. Messer, C. M. Nievergelt, M. J. Arranz, E. Courchesne, K. Pierce, A. R. Muotri, L. M. Iakoucheva, A. Hervas, S. W. Scherer, C. Corsello, and J. Sebat. Paternally inherited cis-regulatory structural variants are associated with autism.* Science*, 360(6386):327–331, 04 2018. PMCID: PMC6449150.*

[171] *E. C. Johnson, R. Border, W. E. Melroy-Greif, C. A. de Leeuw, M. A. Ehringer, and M. C. Keller. No Evidence That Schizophrenia Candidate Genes Are More Associated With Schizophrenia Than Noncandidate Genes.* Biol Psychiatry*, 82(10):702–708, Nov 2017. PMCID: PMC5643230.*

[172] *M. S. Farrell, T. Werge, P. Sklar, M. J. Owen, R. A. Ophoff, M. C. O'Donovan, A. Corvin, S. Cichon, and P. F. Sullivan. Evaluating historical candidate genes for schizophrenia.* Mol Psychiatry*, 20(5):555–62, May 2015. PMCID: PMC4414705.*

[173] *K. E. Samocha, J. A. Kosmicki, K. J. Karczewski, A. H. O'Donnell-Luria, E. Pierce-Hoffman, D. G. MacArthur, B. M. Neale, and M. J. Daly. Regional missense constraint improves variant deleteriousness prediction.* bioRxiv*, 2017.*

[174] *T. J. Hayeck, N. Stong, C. J. Wolock, B. Copeland, S. Kamalakaran, D. B. Goldstein, and A. S. Allen. Improved Pathogenic Variant Localization via a Hierarchical Model of Sub-regional Intolerance.* Am J Hum Genet*, 104(2):299–309, 02 2019. PMCID: PMC6369453.*

[175] *L. Sundaram, H. Gao, S. R. Padigepati, J. F. McRae, Y. Li, J. A. Kosmicki, N. Fritzilas, J. Hakenberg, A. Dutta, J. Shon, J. Xu, S. Batzoglou, X. Li, and K. K.-H. Farh. Predicting the clinical impact of human mutation with deep neural networks.* Nat Genet*, 50(8):1161–1170, 08 2018. PMCID: PMC6237276.*

[176] *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):310–5, Mar 2014. PMCID: PMC3992975.*

[177] *J. Schmidhuber.* Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook*. PhD thesis, Technische Universität München, 1987.*

[178] *S. Thrun and L. Pratt.* Learning to learn*. Springer Science & Business Media, 2012.*

[179] *J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In* International conference on machine learning*, pages 647–655. PMLR, 2014.*

[180] *O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra. Matching networks for one shot learning.* arXiv preprint arXiv:1606.04080*, 2016.*

[181] *A. Baevski, S. Edunov, Y. Liu, L. Zettlemoyer, and M. Auli. Cloze-driven pretraining of self-attention networks.* arXiv preprint arXiv:1903.07785*, 2019.*

[182] *A. Elnaggar, M. Heinzinger, C. Dallago, and B. Rost. End-to-end multitask learning, from protein language to protein features without alignments.* bioRxiv*, page 864405, 2019.*

[183] *W. Kong, R. Somani, Z. Song, S. Kakade, and S. Oh. Meta-learning for mixed linear regression. In* International Conference on Machine Learning*, pages 5394–5404. PMLR, 2020.*

[184] *N. Tripuraneni, M. I. Jordan, and C. Jin. On the theory of transfer learning: The importance of task diversity.* arXiv preprint arXiv:2006.11650*, 2020.*

[185] *N. Tripuraneni, C. Jin, and M. I. Jordan. Provable meta-learning of linear representations.* arXiv preprint arXiv:2002.11684*, 2020.*

[186] *T. T. Cai and H. Wei. Transfer learning for nonparametric classification: Minimax rate and adaptive classifier.* The Annals of Statistics*, 49(1):100–128, 2021.*

[187] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science*, 313(5786):504–507, 2006.*

[188] D. P. Kingma and M. Welling. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114*, 2013.*

[189] Q. Hu and C. S. Greene. Parameter tuning is a key part of dimensionality reduction via deep variational autoencoders for single cell RNA transcriptomics. In BIOCOMPUTING 2019: Proceedings of the Pacific Symposium*, pages 362–373. World Scientific, 2018.*

[190] G. Eraslan, L. M. Simon, M. Mircea, N. S. Mueller, and F. J. Theis. Single-cell RNA-seq denoising using a deep count autoencoder. Nature communications*, 10(1):1–14, 2019.*

[191] L. Lei and W. Fithian. AdaPT: an interactive procedure for multiple testing with side information. Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):649–679, 2018.*

[192] B. Duan, A. Ramdas, and L. Wasserman. Familywise error rate control by interactive unmasking. In International Conference on Machine Learning*, pages 2720–2729. PMLR, 2020.*

[193] P. Chao and W. Fithian. AdaPT-GMM: Powerful and robust covariate-assisted multiple testing. arXiv preprint arXiv:2106.15812*, 2021.*

[194] R. Yurko. Selective inference approaches for augmenting genetic association studies with multi-omics metadata*. PhD thesis, Carnegie Mellon University, 2022.*

[195] W.-P. Lee, A. A. Tucci, M. Conery, Y. Y. Leung, A. B. Kuzma, O. Valladares, Y.-F. Chou, W. Lu, L.-S. Wang, G. D. Schellenberg, and J.-Y. Tzeng. Copy Number Variation Identification on 3,800 Alzheimer's Disease Whole Genome Sequencing Data from the Alzheimer's Disease Sequencing Project. Front Genet*, 12:752390, 2021. PMCID: PMC8599981.*

[196] S. J. Sanders, B. M. Neale, H. Huang, D. M. Werling, J.-Y. An, S. Dong, Whole Genome Sequencing for Psychiatric Disorders (WGSPD), G. Abecasis, P. A. Arguello, J. Blangero, M. Boehnke, M. J. Daly, K. Eggan, D. H. Geschwind, D. C. Glahn, D. B. Goldstein, R. E. Gur, R. E. Handsaker, S. A. McCarroll, R. A. Ophoff, A. Palotie, C. N. Pato, C. Sabatti, M. W. State, A. J. Willsey, S. E. Hyman, A. M. Addington, T. Lehner, and N. B. Freimer. Whole genome sequencing in psychiatric disorders: the WGSPD consortium. Nat Neurosci*, 20(12):1661–1668, 12 2017. PMCID: PMC7785336.*

[197] F. Richter, S. U. Morton, S. W. Kim, A. Kitaygorodsky, L. K. Wasson, K. M. Chen, J. Zhou, H. Qi, N. Patel, S. R. DePalma, M. Parfenov, J. Homsy, J. M. Gorham, K. B. Manheimer, M. Velinder, A. Farrell, G. Marth, E. E. Schadt, J. R. Kaltman, J. W. Newburger, A. Giardini, E. Goldmuntz, M. Brueckner, R. Kim, G. A. Porter, Jr, D. Bernstein, W. K. Chung, D. Srivastava, M. Tristani-Firouzi, O. G. Troyanskaya, D. E. Dickel, Y. Shen, J. G. Seidman, C. E. Seidman, and B. D. Gelb. Genomic analyses implicate noncoding de novo variants in congenital heart disease. Nat Genet*, 52(8):769–777, 08 2020. PMCID: PMC7415662.*

[198] S. U. Morton, A. C. Pereira, D. Quiat, F. Richter, A. Kitaygorodsky, J. Hagen, D. Bernstein, M. Brueckner, E. Goldmuntz, R. W. Kim, R. P. Lifton, G. A. Porter, Jr, M. Tristani-Firouzi, W. K. Chung, A. Roberts, B. D. Gelb, Y. Shen, J. W. Newburger, J. G. Seidman, and C. E. Seidman. Genome-Wide De Novo Variants in Congenital Heart Disease Are Not Associated With Maternal Diabetes or Obesity. Circ Genom Precis Med*, 15(2):e003500, Apr 2022. PMID: 35130025.*