RESEARCH ARTICLE

Quantifying concordant genetic effects of de novo mutations on multiple disorders

Hanmin Guo[1,2], Lin Hou[1,2,3], Yu Shi[4], Sheng Chih Jin[5], Xue Zeng[6,7], Boyang Li[8], Richard P Lifton[6,7], Martina Brueckner[6,9], Hongyu Zhao[6,8,10]*, Qiongshi Lu[11]*

[1]Center for Statistical Science, Tsinghua University, Beijing, China; [2]Department of Industrial Engineering, Tsinghua University, Beijing, China; [3]MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing, China; [4]Yale School of Management, Yale University, New Haven, United States; [5]Department of Genetics, Washington University in St. Louis, St. Louis, United States; [6]Department of Genetics, Yale University, New Haven, United States; [7]Laboratory of Human Genetics and Genomics, Rockefeller University, New York, United States; [8]Department of Biostatistics, Yale School of Public Health, New Haven, United States; [9]Department of Pediatrics, Yale University, New Haven, United States; [10]Program of Computational Biology and Bioinformatics, Yale University, New Haven, United States; [11]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, United States

Abstract Exome sequencing on tens of thousands of parent-proband trios has identified numerous deleterious de novo mutations (DNMs) and implicated risk genes for many disorders. Recent studies have suggested shared genes and pathways are enriched for DNMs across multiple disorders. However, existing analytic strategies only focus on genes that reach statistical significance for multiple disorders and require large trio samples in each study. As a result, these methods are not able to characterize the full landscape of genetic sharing due to polygenicity and incomplete penetrance. In this work, we introduce EncoreDNM, a novel statistical framework to quantify shared genetic effects between two disorders characterized by concordant enrichment of DNMs in the exome. EncoreDNM makes use of exome-wide, summary-level DNM data, including genes that do not reach statistical significance in single-disorder analysis, to evaluate the overall and annotation-partitioned genetic sharing between two disorders. Applying EncoreDNM to DNM data of nine disorders, we identified abundant pairwise enrichment correlations, especially in genes intolerant to pathogenic mutations and genes highly expressed in fetal tissues. These results suggest that EncoreDNM improves current analytic approaches and may have broad applications in DNM studies.

Editor's evaluation

Lu et al. provide a powerful statistical method that measures excess sharing of de novo mutations between pairs of disorders. This method extends the concept of 'genetic correlation' to disorders caused by de-novo mutations, measuring the correlation in excess de-novo mutations in genome-wide genes for different classes of mutations. The authors apply the method to nine disorders including a developmental disorder, autism spectrum disorder, congenital heart disease, schizophrenia, and intellectual disability, finding a statistically significant overlap between 12 pairs of disorders in de novo mutations that cause a loss of gene function. This method will be of interest to researchers working on disorders caused by de-novo mutations.

***For correspondence:** Hongyu.Zhao@yale.edu (HZ); qlu@biostat.wisc.edu (QL)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 16

Preprinted: 14 June 2021
Received: 14 November 2021
Accepted: 01 June 2022
Published: 06 June 2022

Reviewing Editor: Alexander Young, University of California, Los Angeles, United States

Introduction

De novo mutations (DNMs) can be highly deleterious and provide important insights into the genetic cause for disease (***Veltman and Brunner, 2012***). As the cost of sequencing continues to drop, whole-exome sequencing (WES) studies conducted on tens of thousands of family trios have pinpointed numerous risk genes for a variety of disorders (***Lelieveld et al., 2016***; ***Kaplanis et al., 2020***; ***Satterstrom et al., 2020***). In addition, accumulating evidence suggests that risk genes enriched for pathogenic DNMs may be shared by multiple disorders (***Hoischen et al., 2014***; ***Fromer et al., 2014***; ***Homsy et al., 2015***; ***Li et al., 2016***; ***Nguyen et al., 2020***). These shared genes could reveal biological pathways that play prominent roles in disease etiology and shed light on clinically heterogeneous yet genetically related diseases (***Homsy et al., 2015***; ***Li et al., 2016***; ***Nguyen et al., 2020***).

Most efforts to identify shared risk genes directly compare genes that are significantly associated with each disorder (***Nguyen et al., 2017***; ***Willsey et al., 2018***). There have been some successes with this approach in identifying shared genes and pathways (e.g. chromatin modifiers) underlying developmental disorder (DD), autism spectrum disorder (ASD), and congenital heart disease (CHD), thanks to the large trio samples in these studies (***Kaplanis et al., 2020***; ***Satterstrom et al., 2020***; ***Jin et al., 2017***), whereas findings in smaller studies remain suggestive (***Allen et al., 2013***; ***Jin et al., 2020b***). Even in the largest studies to date, statistical power remains moderate for risk genes with weaker effects (***Kaplanis et al., 2020***; ***Howrigan et al., 2020***). It is estimated that more than 1000 haploinsufficient genes contributing to developmental disorder risk have not yet achieved statistical significance in large WES studies (***Kaplanis et al., 2020***). Therefore, analytic approaches that only account for top significant genes cannot capture the full landscape of genetic sharing in multiple disorders. Recently, a Bayesian framework named mTADA was proposed to jointly analyze DNM data of two diseases and improve risk gene mapping (***Nguyen et al., 2020***). Although mTADA produces estimates for the proportion of shared risk genes, the statistical property of these parameter estimates has not been studied. There is a pressing need for powerful, robust, and interpretable methods that quantify concordant DNM association patterns for multiple disorders using exome-wide DNM counts.

Recent advances in estimating genetic correlations using summary data from genome-wide association studies (GWAS) may provide a blueprint for approaching this problem in DNM research (***Zhang et al., 2021a***). Modeling 'omnigenic' associations as independent random effects, linear mixed-effects models leverage genome-wide association profiles to quantify the correlation between additive genetic components of multiple complex traits (***Lee et al., 2012***; ***Bulik-Sullivan et al., 2015***; ***Lu et al., 2017***; ***Ning et al., 2020***). These methods have identified ubiquitous genetic correlations across many human traits and revealed significant and robust genetic correlations that could not be inferred from significant GWAS associations alone (***Shi et al., 2017***; ***Brainstorm, 2018***; ***Guo et al., 2021***; ***Zhang et al., 2021b***).

Here, we introduce EncoreDNM (**En**richment **cor**relation **e**stimator for **D**e **N**ovo **M**utations), a novel statistical framework that leverages exome-wide DNM counts, including genes that do not reach exome-wide statistical significance in single-disorder analysis, to estimate concordant DNM associations between disorders. EncoreDNM uses a generalized linear mixed-effects model to quantify the occurrence of DNMs while accounting for de novo mutability of each gene and technical inconsistencies between studies. We demonstrate the performance of EncoreDNM through extensive simulations and analyses of DNM data of nine disorders.

Results

Method overview

DNM counts in the exome deviate from the null (i.e. expected counts based on de novo mutability) when mutations play a role in disease etiology. Disease risk genes will show enrichment for deleterious DNMs in probands and non-risk genes may be slightly depleted for DNM counts. Our goal is to estimate the correlation of such deviation between two disorders, which we refer to as the DNM enrichment correlation. More specifically, we use a pair of mixed-effects Poisson regression models (***Munkin and Trivedi, 1999***) to quantify the occurrence of DNMs in two studies.

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} \sim \text{Poisson}\left(\begin{bmatrix} \lambda_{i1} \\ \lambda_{i2} \end{bmatrix}\right),$$

$$\log\left(\begin{bmatrix} \lambda_{i1} \\ \lambda_{i2} \end{bmatrix}\right) = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \log\left(\begin{bmatrix} 2N_1 m_i \\ 2N_2 m_i \end{bmatrix}\right) + \begin{bmatrix} \phi_{i1} \\ \phi_{i2} \end{bmatrix},$$

$$\begin{bmatrix} \phi_{i1} \\ \phi_{i2} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right).$$

Here, Y_{i1}, Y_{i2} are the DNM counts for the i-th gene and N_1, N_2 are the number of parent-proband trios in two studies, respectively. The log Poisson rates of DNM occurrence are decomposed into three components: the elevation component, the background component, and the deviation component. The elevation component β_k $(k = 1, 2)$ is a fixed effect term adjusting for systematic, exome-wide bias in DNM counts. One example of such bias is the batch effect caused by different sequencing and variant calling pipelines in two studies. The elevation parameter β_k tends to be positive when DNMs are over-called with higher sensitivity and negative when DNMs are under-called with higher specificity (***Wei et al., 2015***). The background component $\log\left(2N_k m_i\right)$ is a gene-specific fixed effect that reflects the expected mutation counts determined by the genomic sequence context under the null (***Samocha et al., 2014***). m_i is the de novo mutability for the i-th gene, and $2N_1 m_i$ and $2N_2 m_i$ are the expected DNM counts in the i-th gene under the null in two studies. The deviation component ϕ_{ik} is a gene-specific random effect that quantifies the deviation of DNM profile from what is expected under the null (i.e. no risk genes for the disorder). ϕ_{i1} and ϕ_{i2} follow a multivariate normal distribution with dispersion parameters σ_1 and σ_2 and a correlation ρ. A larger value of the dispersion parameter σ_k indicates a more substantial deviation from the null. That is, DNM counts show strong enrichment in some genes and depletion in other genes compared to the expectation based on de novo mutability. A smaller value of σ_k suggests that the DNM count data is well in line with what is expected based on

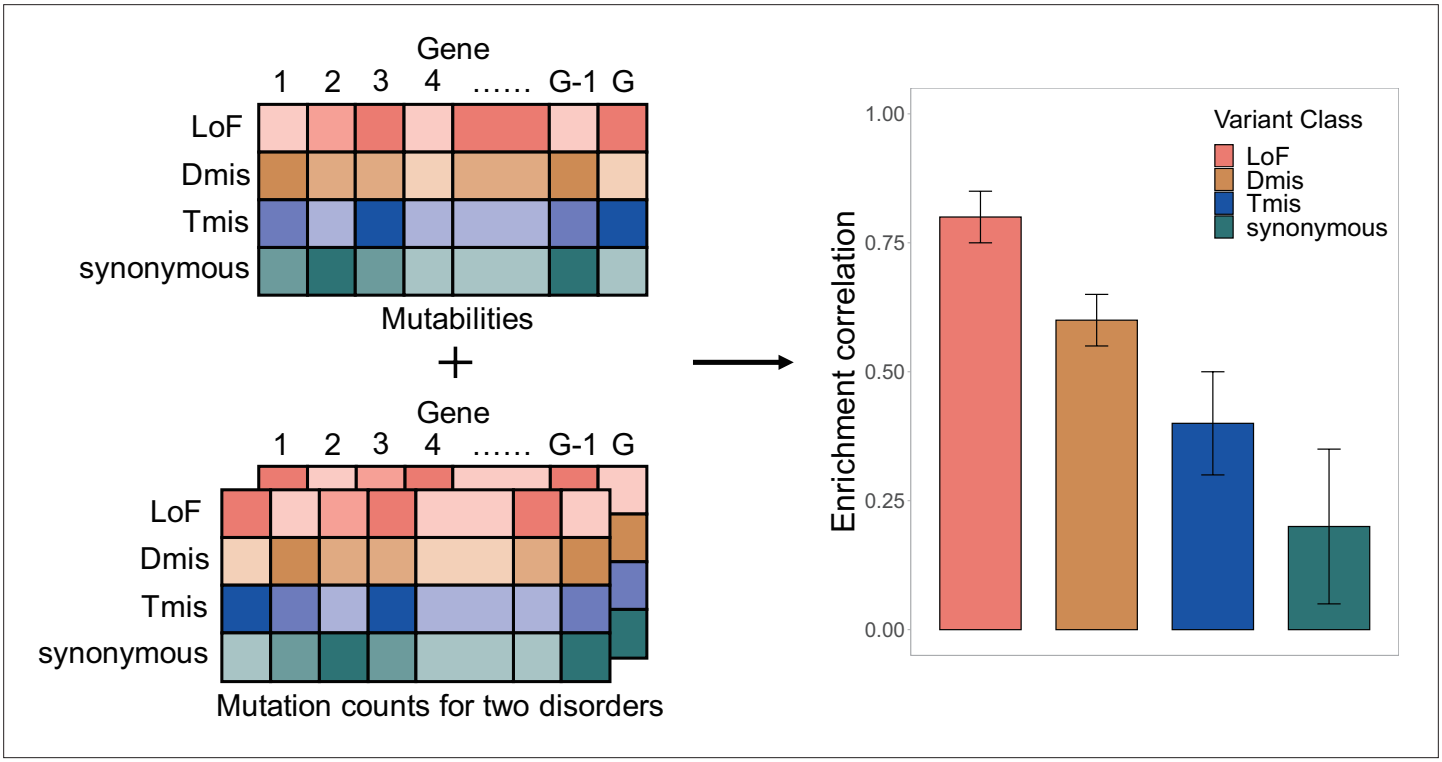

Figure 1. EncoreDNM workflow. The inputs of EncoreDNM are de novo mutability of each gene and exome-wide, annotated DNM counts from two studies. We fit a mixed-effects Poisson model to estimate the DNM enrichment correlation between two disorders for each variant class.

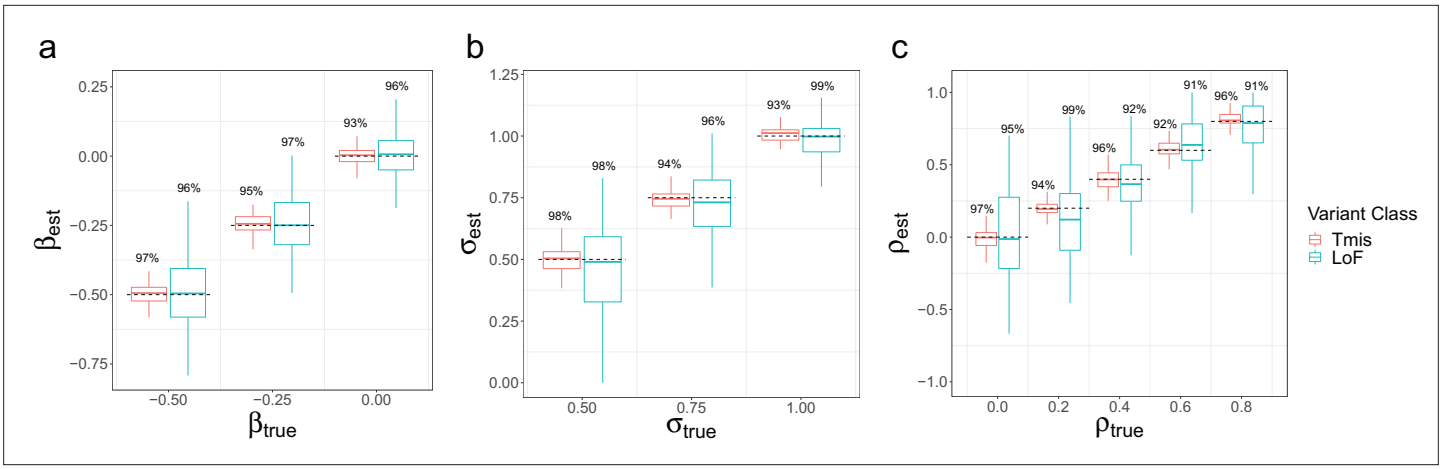

Figure 2. Parameter estimation results of EncoreDNM. (**a**) Boxplot of β estimates in single-trait analysis with σ fixed at 0.75. (**b**) Boxplot of σ estimates in single-trait analysis with β fixed at –0.25. (**c**) Boxplot of ρ estimates in cross-trait analysis with β and σ fixed at –0.25 and 0.75. True parameter values are marked by dashed lines. The number above each box represents the coverage rate of 95% Wald confidence intervals. Each simulation setting was repeated 100 times.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Estimation results of elevation parameter $\boldsymbol{\beta}$ under a mixed-effects Poisson regression model.

Figure supplement 2. Estimation results of dispersion parameter $\boldsymbol{\sigma}$ under a mixed-effects Poisson regression model.

the null model. DNM enrichment correlation is denoted by ρ and is our main parameter of interest. It quantifies the concordance of DNM burden in two disorders.

Parameters in this model can be estimated using a Monte Carlo maximum likelihood estimation (MLE) procedure. Standard errors of the estimates are obtained through inversion of the observed Fisher information matrix. In practice, we use annotated DNM data as input and fit mixed-effects Poisson models for each variant class separately: loss of function (LoF), deleterious missense (Dmis, defined as MetaSVM-deleterious), tolerable missense (Tmis, defined as MetaSVM-tolerable), and synonymous (***Figure 1***). More details about model setup and parameter estimation are discussed in Materials and methods.

Simulation results

We conducted simulations to assess the parameter estimation performance of EncoreDNM in various settings. We focused on two variant classes, that is, Tmis and LoF variants, since they have the highest and lowest median mutabilities in the exome. We used EncoreDNM to estimate the elevation parameter β, dispersion parameter σ, and enrichment correlation ρ (Materials and methods). Under various parameter settings, EncoreDNM always provided unbiased estimation of the parameters (***Figure 2*** and ***Figure 2—figure supplements 1–2***). Furthermore, the 95% Wald confidence intervals achieved coverage rates close to 95% under all simulation settings, demonstrating the effectiveness of EncoreDNM to provide accurate statistical inference.

Next, we compared the performance of EncoreDNM with mTADA (***Nguyen et al., 2020***), a Bayesian framework that estimates the proportion of shared risk genes for two disorders. First, we simulated DNM data under the mixed-effects Poisson model. We evaluated two methods across a range of combinations of elevation parameter, dispersion parameter, and sample size for two disorders. The false positive rates for our method were well-calibrated in all parameter settings, but mTADA produced false positive findings when the observed DNM counts were relatively small (e.g. due to reduced elevation or dispersion parameters or a lower sample size; ***Figure 3a***). We also assessed the statistical power of two approaches under a baseline setting where false positives for both methods were controlled. As enrichment correlation increased, EncoreDNM achieved universally greater statistical power compared to mTADA (***Figure 3b***).

To ensure a fair comparison, we also considered a mis-specified model setting where we randomly distributed the total DNM counts for each disorder into all genes with an enrichment in causal genes (Materials and methods). EncoreDNM showed well-controlled false positive rate across all simulation

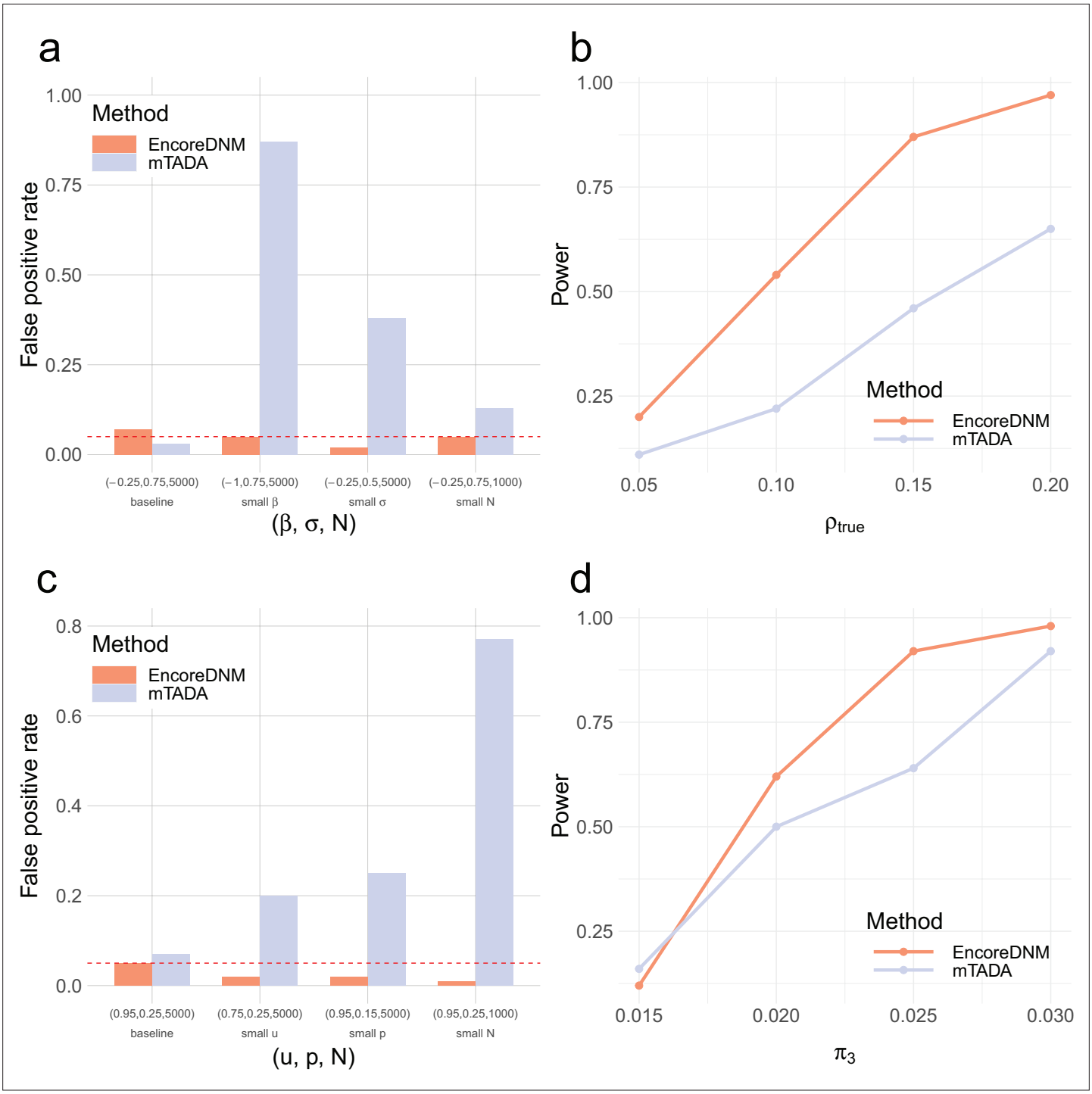

Figure 3. Comparison of EncoreDNM and mTADA. (**a**) False positive rates under a mixed-effects Poisson regression model. (**b**) Statistical power of two methods under a mixed-effects Poisson regression model as the enrichment correlation increases. Parameters (β, σ, N) were fixed at (–0.25, 0.75, 5000) for both disorders. (**c**) False positive rates under a multinomial model. (**d**) Statistical power under a multinomial model with varying proportion of shared causal genes. Parameters (u, p, N) were fixed at (0.95, 0.25, 5000) for both disorders. Each simulation setting was repeated 100 times.

settings, whereas severe inflation of false positives arose for mTADA when the total mutation count, the proportion of probands that can be explained by DNMs, or the sample size were small (***Figure 3c***). Furthermore, we compared the statistical power of two methods under this model in a baseline setting where false positive rate was controlled. EncoreDNM showed higher statistical power compared to mTADA as the fraction of shared causal genes increased (***Figure 3d***).

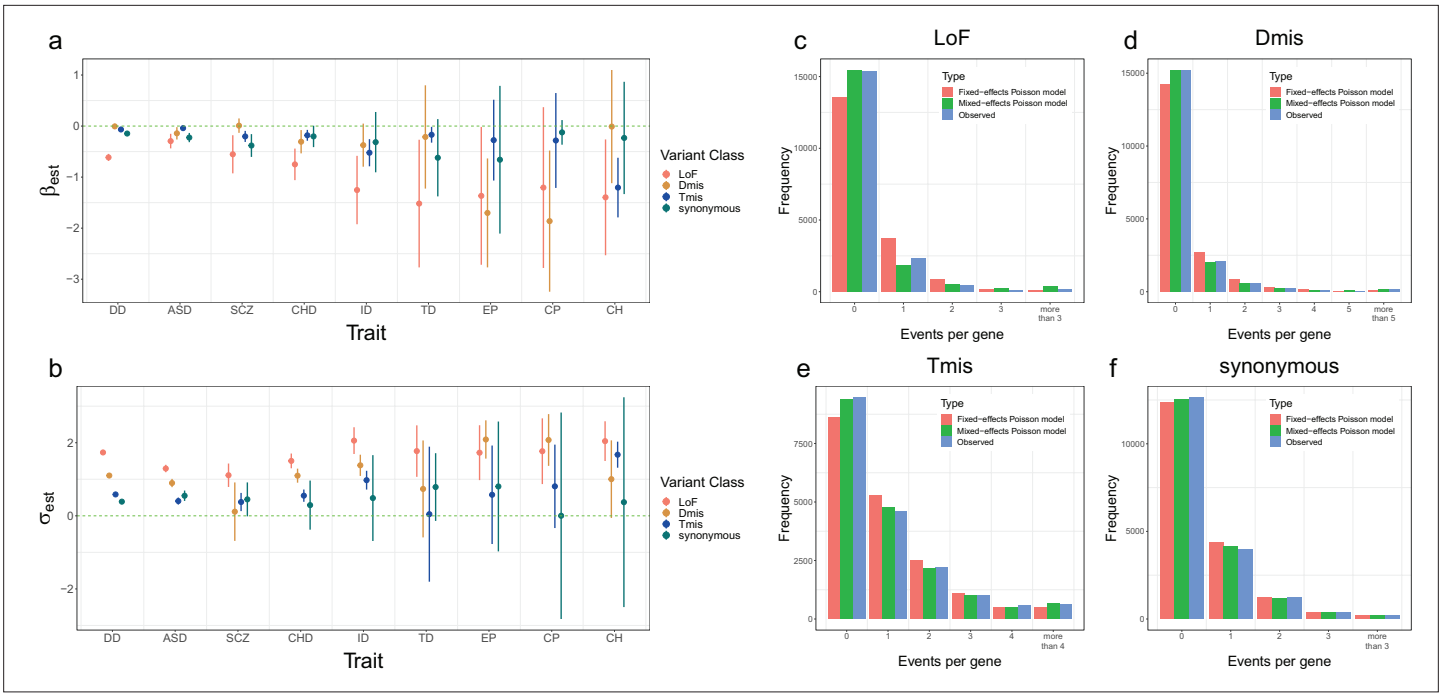

Figure 4. Model fitting results for nine disorders. (**a, b**) Estimation results of β and σ for nine disorders and four variant classes. Error bars represent 1.96*standard errors. Sample sizes of DNM datasets for each disorder are provided in ***Supplementary file 1***-STable 1. (**c–f**) Distribution of DNM events per gene in four variant classes for developmental disorder. Red and green bars represent the expected frequency of genes under the fixed-effects and mixed-effects Poisson regression models, respectively. Blue bars represent the observed frequency of genes.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Likelihood ratio test shows significantly improved goodness of fit of the mixed-effects Poisson model compared to a fixed-effects model without the deviation component.

Pervasive enrichment correlation of damaging DNMs among developmental disorders

We applied EncoreDNM to DNM data of nine disorders (***Supplementary file 1***-STable 1; Materials and methods): developmental disorder (n=31,058; number of trios; ***Kaplanis et al., 2020***), autism spectrum disorder (n=6430; ***Satterstrom et al., 2020***), schizophrenia (SCZ; n=2772; ***Howrigan et al., 2020***), congenital heart disease (n=2645; ***Jin et al., 2017***), intellectual disability (ID; n=820; ***Lelieveld et al., 2016***), Tourette disorder (TD; n=484) ***Willsey et al., 2017***, epileptic encephalopathies (EP; n=264; ***Allen et al., 2013***), cerebral palsy (CP; n=250; ***Jin et al., 2020b***), and congenital hydrocephalus (CH; n=232; ***Jin et al., 2020a***). In addition, we also included 1789 trios comprising healthy parents and unaffected siblings of autism probands as controls (***Krumm et al., 2015***).

We first performed single-trait analysis under the mixed-effects Poisson model for each disorder. The estimated elevation parameters (i.e. β) were negative for almost all disorders and variant classes (***Figure 4a***), with LoF variants showing particularly lower parameter estimates. This may be explained by more stringent quality control in LoF variant calling (***Jin et al., 2017***) and potential survival bias (***Lek et al., 2016***). It is also consistent with a depletion of LoF DNMs in healthy control trios (***Homsy et al., 2015***). The dispersion parameter estimates (i.e. σ) were higher for LoF variants than other variant classes (***Figure 4b***), which is consistent with our expectation that LoF variants have stronger effects on disease risk and should show a larger deviation from the null mutation rate in disease probands. We also compared the goodness of fit of our proposed mixed-effects Poisson model to a simpler fixed-effects model without the deviation component (Materials and methods). The expected distribution of recurrent DNM counts showed substantial and statistically significant improvement under the mixed-effects Poisson model (***Figure 4c–f***, ***Figure 4—figure supplement 1***, and ***Supplementary file 1***-STable 2).

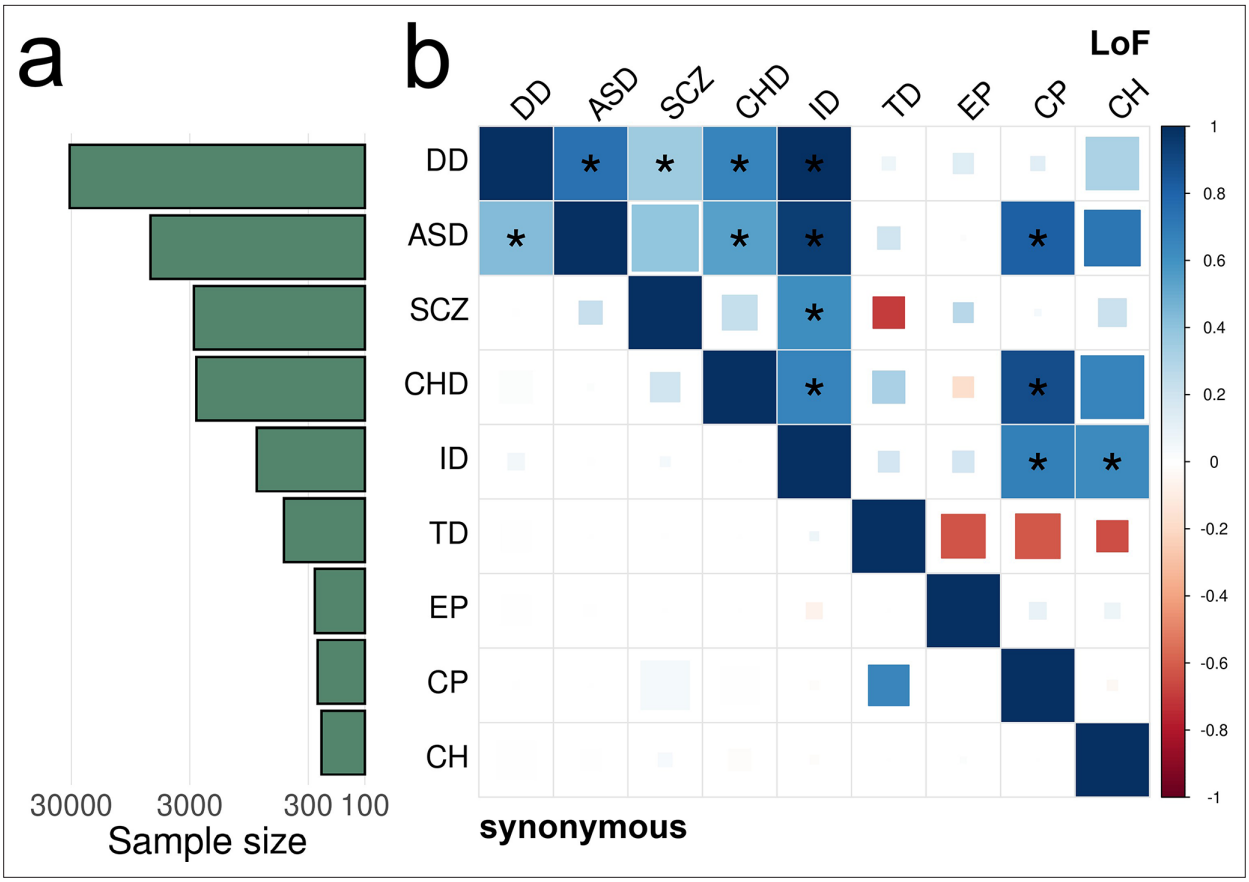

Figure 5. EncoreDNM identifies pervasive enrichment correlations of damaging DNMs among nine disorders. (**a**) Shows sample size (for example, number of trios) for each disease. X-axis denotes sample size on the log scale. (**b**) Heatmap of enrichment correlations for LoF (upper triangle) and synonymous (lower triangle) DNMs among nine disorders. Larger squares represent more significant p-values, and deeper color represents stronger correlations. Significant correlations (FDR <0.05) are shown as full-sized squares marked by asterisks.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. DNM enrichment correlations of nine disorders based on Dmis and Tmis variants.

Figure supplement 2. DNM enrichment correlations between nine disorders and controls.

Figure supplement 3. Number of significant correlations identified for each disorder is proportional to its sample size.

Figure supplement 4. Lollipop plot for LoF DNMs in *CTNNB1*.

Figure supplement 5. Lollipop plot for LoF DNMs in *FBXO11*.

Figure supplement 6. DNM genetic sharing in nine disorders estimated for LoF, Dmis, Tmis, and synonymous DNMs using mTADA.

Figure supplement 7. DNM genetic sharing in nine disorders and controls identified by mTADA.

Figure supplement 8. Comparison of GWAS- and DNM-based estimation of genetic sharing among five disorders.

Figure supplement 9. Group-wise jackknife method and inversion of Fisher information matrix method produced similar standard error estimates for LoF variants.

Next, we estimated pairwise DNM enrichment correlations for 9 disorders. In total, we identified 25 pairs of disorders with significant correlations at a false discovery rate (FDR) cutoff of 0.05 (***Figure 5*** and ***Figure 5—figure supplement 1***), including 12 significant correlations for LoF variants, 7 for Dmis variants, 5 for Tmis variants, and only 1 significant correlation for synonymous variants. Notably, all significant correlations are positive (***Supplementary file 1***-STable 3). No significant correlation was identified between any disorder and healthy controls (***Figure 5—figure supplement 2***). This is consistent with our expectation, since DNMs in the control groups will distribute proportionally according to the de novo mutability without showing enrichment in certain genes. The number of identified significant correlations for each disorder was proportional to the sample size in each study (Spearman correlation = 0.70) with controls being a notable outlier (***Figure 5—figure supplement 3***).

We identified highly concordant and significant LoF DNM enrichment among developmental disorder, autism, intellectual disability, and congenital heart disease, which is consistent with previous reports (***Li et al., 2016***; ***Nguyen et al., 2020***; ***Nguyen et al., 2017***; ***Hormozdiari et al., 2015***). Schizophrenia shows highly significant LoF correlations with developmental disorder (p=2.0e-3) and intellectual disability (3.7e-5). The positive enrichment correlation between autism and cerebral palsy in LoF variants (ρ=0.81, p=3.3e-3) is consistent with their co-occurrence (***Christensen et al., 2014***). The high enrichment correlation between intellectual disability and cerebral palsy in LoF variants (ρ=0.68, p=1.0e-4) is consistent with the associations between intellectual disability and motor or non-motor abnormalities caused by cerebral palsy (***Reid et al., 2018***). A previous study also suggested significant genetic sharing of intellectual disability and cerebral palsy by overlapping genes harboring rare damaging variants (***Jin et al., 2020b***). Here, we obtained consistent results after accounting for de novo mutabilities and potential confounding bias.

Some significant correlations identified in our analysis are consistent with phenotypic associations in epidemiological studies, but have not been reported using genetic data to the extent of our knowledge. For example, the LoF enrichment correlation between congenital heart disease and cerebral palsy (ρ=0.88, p=1.7e-3) is consistent with findings that reduced supply of oxygenated blood in fetal brain due to cardiac malformations may be a risk factor for cerebral palsy (***Garne et al., 2008***). The enrichment correlation between intellectual disability and congenital hydrocephalus in LoF variants (ρ=0.63, p=2.4e-3) is consistent with lower intellectual performance in a proportion of children with congenital hydrocephalus (***Lumenta and Skotarczak, 1995***).

Genes showing pathogenic DNMs in multiple disorders may shed light on the mechanisms underlying enrichment correlations (***Supplementary file 1***-STable 4). We identified five genes (*CTNNB1*, *NBEA*, *POGZ*, *SPRED2*, and *KMT2C*) with LoF DNMs in five different disorders and 21 genes had LoF DNMs in four disorders (***Supplementary file 1***-STable 5). These 26 genes with LoF variants in at least four disorders were significantly enriched for 63 gene ontology (GO) terms with FDR <0.05 (***Supplementary file 1***-STable 6). Chromatin organization (p=7.8e-11), nucleoplasm (p=2.8e-10), chromosome organization (p=6.8e-10), histone methyltransferase complex (p=1.4e-9), and positive regulation of gene expression (p=2.2e-9) were the most significantly enriched GO terms. One notable example consistently included in these gene sets is *CTNNB1* (***Figure 5—figure supplement 4***). It encodes β-catenin, is one of the only two genes reaching genome-wide significance in a recent WES study for cerebral palsy (***Jin et al., 2020b***), and also harbors multiple LoF variants in developmental disorder, intellectual disability, autism, and congenital heart disease. It is a fundamental component of the canonical Wnt signaling pathway which is known to confer genetic risk for autism (***O'Roak et al., 2012***). Genes with recurrent damaging DNMs in multiple disorders also revealed shared biological function across these disorders (***Rees et al., 2021***). We identified 30 recurrent cross-disorder LoF mutations that were not recurrent in developmental disorder alone (***Supplementary file 1***). *FBXO11*, encoding the F-box only protein 31, shows two recurrent p.Ser831fs LoF variants in autism and congenital hydrocephalus (***Figure 5—figure supplement 5***; p=1.9e-3; Materials and methods). The F-box protein constitutes a substrate-recognition component of the SCF (SKP1-cullin-F-box) complex, an E3-ubiquitin ligase complex responsible for ubiquitination and proteasomal degradation (***Cardozo and Pagano, 2004***). DNMs in *FBXO11* have been previously implicated in severe intellectual disability individuals with autistic behavior problem (***Jansen et al., 2019***) and neurodevelopmental disorder (***Gregor et al., 2018***).

For comparison, we also applied mTADA to the same nine disorders and control trios. In total, mTADA identified 117 disorder pairs with significant genetic sharings at an FDR cutoff of 0.05 (***Supplementary file 1***-STable 8 and ***Figure 5—figure supplement 6***). Notably, we identified significant synonymous DNM correlations for all 36 disorder pairs and between all disorders and healthy controls (***Figure 5—figure supplement 7***). These results are consistent with the simulation results and suggest a substantially inflated false positive rate in mTADA.

Partitioning DNM enrichment correlation by gene set

To gain biological insights into the shared genetic architecture of nine disorders, we repeated EncoreDNM correlation analysis in several gene sets. First, we defined genes with high/low probability of intolerance to LoF variants using pLI scores (***Karczewski et al., 2020***), and identified genes with high/low brain expression (HBE/LBE) (***Werling et al., 2020***; Materials and methods; ***Supplementary***

file 1-STable 9). We identified 11 and 12 disorder pairs showing significant enrichment correlations for LoF DNMs in high-pLI genes and HBE genes, respectively (***Figure 6a–b***). We observed fewer significant correlations for Dmis and Tmis variants in these gene sets (***Figure 6—figure supplements 1–2***). All identified significant correlations were positive (***Supplementary file 1***-STables 10 -11). No significant correlations were identified for synonymous variants (***Figure 6—figure supplements 1–2***) or between disorders and controls (***Figure 6—figure supplements 3–4***).

We observed a clear enrichment of significant correlations in disease-relevant gene sets. Overall, high-pLI genes showed substantially stronger correlations across disorders than genes with low pLI (one-sided Kolmogorov-Smirnov test; p=2.3e-6). Similarly, enrichment correlations were stronger in HBE genes than in LBE genes (p=8.8e-7). Among the 11 disorder pairs showing significant enrichment correlations in high-pLI genes, two pairs, that is, autism-schizophrenia (ρ=0.68, p=2.4e-3) and developmental disorder-congenital hydrocephalus (ρ=0.43, p=1.5e-3), were not identified in the exome-wide analysis. We also identified four novel disorder pairs with significant correlations in HBE genes, including developmental disorder-cerebral palsy (ρ=0.80, p=9.5e-5), developmental disorder-congenital hydrocephalus (ρ=0.67, p=1.4e-3), autism-congenital hydrocephalus (ρ=0.82, p=4.7e-4), and schizophrenia-epileptic encephalopathies (ρ=0.66, p=2.0e-3). These novel enrichment correlations are consistent with known comorbidities between these disorders (***Kielinen et al., 2004***; ***Kilincaslan and Mukaddes, 2009***) and findings based on significant risk genes (***Li et al., 2016***; ***Jin et al., 2020a***; ***Kume et al., 1998***; ***Cao and Wu, 2015***).

Furthermore, we estimated DNM enrichment correlations in genes with high/low expression in mouse developing heart (HHE/LHE) (***Homsy et al., 2015***; Materials and methods; ***Supplementary file 1***-STable 9). We identified 9 significant enrichment correlations for LoF variants in HHE genes (***Figure 6c***). Strength of enrichment correlations did not show a significant difference between HHE and LHE genes (p=0.846), possibly due to a lack of cardiac disorders in our analysis. Finally, we estimated enrichment correlations between congenital heart disease and other disorders in known pathways for congenital heart disease (***Zaidi and Brueckner, 2017***; Materials and methods**;** ***Supplementary file 1***-STable 9). We identified five significant correlations for LoF variants (***Figure 6d***), including a novel correlation between congenital heart disease and Tourette disorder (ρ=0.93, p=3.3e-9). Of note, arrhythmia caused by congenital heart disease is a known risk factor for Tourette disorder (***Gulisano et al., 2011***). In these analyses, all significant enrichment correlations were positive (***Supplementary file 1***) and other variant classes showed generally weaker correlations than LoF variants (***Figure 6—figure supplements 5–6***). We did not observe significant correlations in these gene sets between disorders and controls (***Figure 6—figure supplements 7–8***).

Discussion

In this paper, we introduced EncoreDNM, a novel statistical framework to quantify correlated DNM enrichment between two disorders. Through extensive simulations and analyses of DNM data for nine disorders, we demonstrated that our proposed mixed-effects Poisson regression model provides unbiased parameter estimates, shows well-controlled false positive rate, and is robust to exome-wide technical biases. Leveraging exome-wide DNM counts and genomic context-based mutability data, EncoreDNM achieves superior fit for real DNM datasets compared to simpler models and provides statistically powerful and computationally efficient estimation of DNM enrichment correlation. Further, EncoreDNM can quantify concordant genetic effects for user-defined variant classes within pre-specified gene sets, thus is suitable for exploring diverse types of hypotheses and can provide crucial biological insights into the shared genetic etiology in multiple disorders. In comparison, the Bayesian approach implemented in mTADA can produce false positives findings, especially when the DNM count is low, possibility due to the overestimated proportion of risk genes. We still observed inflation in false positive rates under a more stringent significance cutoff or using posterior probability threshold strategy (***Supplementary file 1***-STables 14-17).

Multi-trait analyses of GWAS data have revealed shared genetic architecture among many neuropsychiatric traits (***Brainstorm, 2018***; ***Lee et al., 2013***; ***Gratten et al., 2014***; ***Abdellaoui and Verweij, 2021***). These findings have led to the identification of pleiotropic variants, genes, and hub genomic regions underlying many traits and have revealed multiple psychopathological factors jointly affecting human neurological phenotypes (***Lee, 2019***; ***Wang et al., 2015***). Although emerging evidence suggests that causal DNMs underlying several disorders with well-powered studies (e.g. congenital

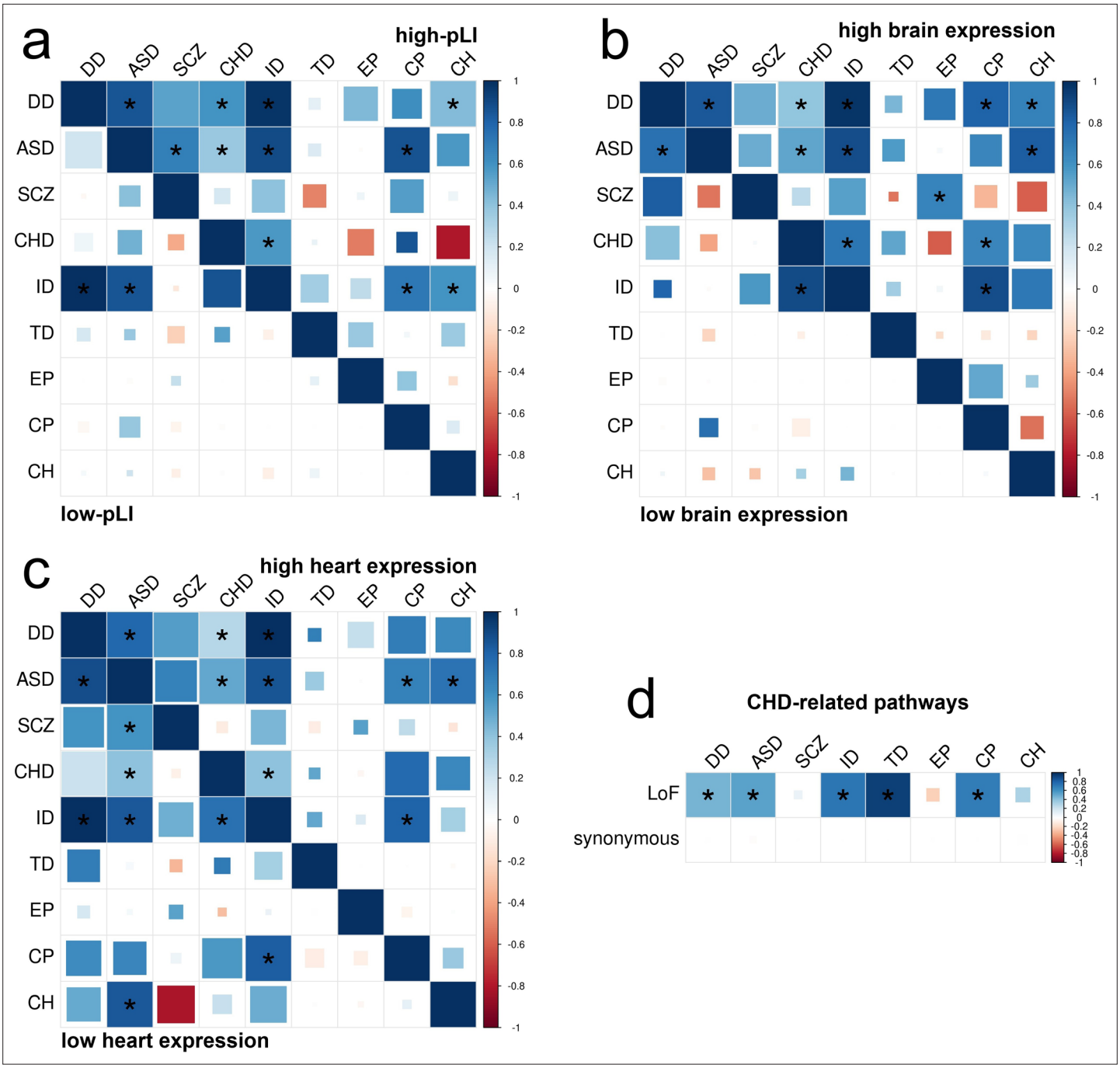

Figure 6. DNM enrichment correlations in disease-relevant gene sets. (**a**) Enrichment correlations in high-pLI genes (upper triangle) and low-pLI genes (lower triangle) for LoF variants. Here, pLI is the probability of being loss-of-function intolerant (see Materials and methods). (**b**) Enrichment correlations in HBE genes (upper triangle) and LBE genes (lower triangle) for LoF variants. (**c**) Enrichment correlations in HHE genes (upper triangle) and LHE genes (lower triangle) for LoF variants. (**d**) Enrichment correlations in CHD-related pathways for LoF and synonymous variants. Larger squares represent more significant p-values, and deeper color represents stronger correlations. Significant correlations (FDR <0.05) are shown as full-sized squares marked by asterisks.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. DNM enrichment correlations in high-pLI genes (upper triangle) and low-pLI genes (lower triangle) for Dmis, Tmis, and synonymous variants.

Figure supplement 2. DNM enrichment correlations in HBE genes (upper triangle) and LBE genes (lower triangle) for Dmis, Tmis, and synonymous variants.

Figure 6 continued on next page

Figure 6 continued

Figure supplement 3. DNM enrichment correlations between nine disorders and controls in high-pLI and low-pLI gene sets.

Figure supplement 4. DNM enrichment correlations between nine disorders and controls in HBE and LBE genes.

Figure supplement 5. DNM enrichment correlations in HHE genes (upper triangle) and LHE genes (lower triangle) for Dmis, Tmis, and synonymous variants.

Figure supplement 6. DNM enrichment correlations in CHD-related pathways for Dmis and Tmis variants.

Figure supplement 7. DNM enrichment correlations between nine disorders and controls in HHE and LHE gene sets.

Figure supplement 8. DNM enrichment correlations between CHD and controls in CHD-related pathways.

heart disease and neurodevelopmental disorders; ***Homsy et al., 2015***) may be shared, our understanding of the extent and the mechanism underlying such sharing remains incomplete. Applied to DNM data for nine disorders, EncoreDNM identified pervasive enrichment correlations of DNMs. We observed particularly strong correlations in pathogenic variant classes (e.g. LoF and Dmis variants) and disease-relevant genes (e.g. genes with high pLI and genes highly expressed in relevant tissues). Genes underlying these correlations were significantly enriched in pathways involved in chromatin organization and modification and gene expression regulation. The DNM correlations were substantially attenuated in genes with lower expression and genes with frequent occurrences of LoF variants in the population. A similar attenuation was observed in less pathogenic variant classes (e.g., synonymous variants). Further, no significant correlations were identified between any disorder and healthy controls. We also compared DNM enrichment correlations of five disorders with genetic correlations estimated from GWAS summary statistics (***Supplementary file 1***-STable 18). We had consistent findings from GWAS and DNM data (Spearman correlation = 0.70; ***Figure 5—figure supplement 8*** and ***Supplementary file 1***-STable 19). These results lay the groundwork for future investigations of pleiotropic mechanisms of DNMs.

Our study has some limitations. First, EncoreDNM assumes probands from different input studies to be independent. In rare cases when two studies have overlapping proband samples, enrichment correlation estimates may be inflated and must be interpreted with caution. Second, genetic correlation methods based on GWAS summary data provided key motivations for the mixed-effects Poisson regression model in our study. Built upon genetic correlations, a plethora of methods have been developed in the GWAS literature to jointly model more than two GWAS (***Turley et al., 2018***), identify and quantify common factors underlying multiple traits (***Grotzinger et al., 2019***; ***Grotzinger et al., 2020***), estimate causal effects among different traits (***Pickrell et al., 2016***), and identify pleiotropic genomic regions through hypothesis-free scans (***Guo et al., 2021***). Future directions of EncoreDNM include using enrichment correlation to improve gene discovery, learning the directional effects and the causal structure underlying multiple disorders, and dynamically searching for gene sets and annotation classes with shared genetic effects without pre-specifying the hypothesis.

Taken together, we provide a new analytic approach to an important problem in DNM studies. We believe EncoreDNM improves the statistical rigor in multi-disorder DNM modeling and opens up many interesting future directions in both method development and follow-up analyses in WES studies. As trio sample size in WES studies continues to grow, EncoreDNM will have broad applications and can greatly benefit DNM research.

Materials and methods

Statistical model

For a single study, we assume that DNM counts in a given variant class (for example, synonymous variants) follow a mixed-effects Poisson model:

$$Y_i \sim \text{Poisson}\left(\lambda_i\right),$$

$$\log\left(\lambda_i\right) = \beta + \log\left(2Nm_i\right) + \phi_i,$$

$$\phi_i \sim \text{N}\left(0, \sigma^2\right), \quad \text{for } i = 1, \ldots, G,$$

where Y_i is the DNM count in the i-th gene, N is the number of trios, m_i is the de novo mutability for the i-th gene (for example, mutation rate per chromosome per generation) which is known a priori (***Samocha et al., 2014***), and G is the total number of genes in the study. The elevation parameter β quantifies the global elevation of mutation rate compared to mutability estimates based on genomic sequence alone. Gene-specific deviation from expected DNM rate is quantified by random effect ϕ_i with a dispersion parameter σ. Here, the ϕ_i are assumed to be independent across different genes, in which case the observed DNM counts of different genes are independent. There is no constraint on the value of β, and the dispersion parameter σ can be any positive value.

Next, we describe how we expand this model to quantify the shared genetics of two disorders. We adopt a flexible Poisson-lognormal mixture framework that can accommodate both overdispersion and correlation (***Munkin and Trivedi, 1999***). We assume DNM counts in a given variant class for two diseases follow:

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} \sim \text{Poisson}\left(\begin{bmatrix} \lambda_{i1} \\ \lambda_{i2} \end{bmatrix}\right),$$

$$\log\left(\begin{bmatrix} \lambda_{i1} \\ \lambda_{i2} \end{bmatrix}\right) = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \log\left(\begin{bmatrix} 2N_1 m_i \\ 2N_2 m_i \end{bmatrix}\right) + \begin{bmatrix} \phi_{i1} \\ \phi_{i2} \end{bmatrix},$$

$$\begin{bmatrix} \phi_{i1} \\ \phi_{i2} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right),$$

where Y_{i1}, Y_{i2} are the DNM counts for the i-th gene and N_1, N_2 are the trio sizes in two studies, respectively. Similar to the single-trait model, m_i is the mutability for the i-th gene. β_1, β_2 are the elevation parameters, and ϕ_{i1}, ϕ_{i2} are the gene-specific random effects with dispersion parameters σ_1, σ_2 , for two disorders respectively. ρ is the enrichment correlation which quantifies the concordance of the gene-specific DNM burden between two disorders. Here, $\beta_1, \beta_2, \sigma_1, \sigma_2, \rho$ are unknown parameters. The gene specific effects for two disorders are assumed to be independent for different genes. We also assume that there is no shared sample for two disorders, in which case Y_{i1} is independent with Y_{i2} given $\begin{bmatrix} \lambda_{i1} \\ \lambda_{i2} \end{bmatrix}$.

Parameter estimation

We implement an MLE procedure to estimate unknown parameters. For single-trait analysis, the log-likelihood function can be expressed as follows:

$$l\left(\beta, \sigma | \boldsymbol{Y}\right) = \sum_{i=1}^{G} \log\left[\int \exp\left(-\lambda_i\right) \lambda_i^{Y_i} * f\left(\phi_i\right) d\phi_i\right] + C,$$

where $\boldsymbol{Y} = \left[Y_1, \ldots, Y_G\right]^T$, $\lambda_i = 2Nm_i \exp\left(\beta + \phi_i\right)$, $C = -\sum_{i=1}^{G} \log\left(Y_i!\right)$, and $f\left(\phi_i\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\phi_i^2}{2\sigma^2}\right)$. Note that there is no closed form for the integral in the log-likelihood function. Therefore, we use Monte Carlo integration to evaluate the log-likelihood function. Let $\phi_{ij} = \sigma\xi_{ij}$, where the ξ_{ij} are independently and identically distributed random variables following a standard normal distribution. We have

$$l\left(\beta, \sigma | \boldsymbol{Y}\right) \approx l'\left(\beta, \sigma | \boldsymbol{Y}\right) = \sum_{i=1}^{G} \log\left[\sum_{j=1}^{M} \exp\left(-\lambda_{ij}\right) \lambda_{ij}^{Y_i}\right] + C,$$

where $\lambda_{ij} = 2Nm_i \exp\left(\beta + \sigma\xi_{ij}\right)$, and M is the Monte Carlo sample size which is set to be 1,000. Then, we could obtain the MLE of β, σ through maximization of $l'\left(\beta, \sigma | \boldsymbol{Y}\right)$. We obtain the standard error of the MLE through inversion of the observed Fisher information matrix. However, when the DNM count is small, the Fisher information may be non-invertible and the parameter vector is not numerically identifiable. In this case, we employ group-wise jackknife using 100 randomly partitioned gene groups to obtain standard errors for parameter estimates. This approach produces consistent standard errors compared to the Fisher information approach (***Figure 5—figure supplement 9***).

The estimation procedure can be generalized to multi-trait analysis. Log-likelihood function can be expressed as follows:

$$l\left(\beta_1,\beta_2,\sigma_1,\sigma_2,\rho|\boldsymbol{Y}_1,\boldsymbol{Y}_2\right)=\sum_{i=1}^{G}\log\left[\int\exp\left(-\lambda_{i1}-\lambda_{i2}\right)\lambda_{i1}^{Y_{i1}}\lambda_{i2}^{Y_{i2}}*f\left(\phi_{i1},\phi_{i2}\right)d\phi_{i1}d\phi_{i2}\right]+C,$$

where $\boldsymbol{Y}_1=\left[Y_{11},\ldots,Y_{G1}\right]^T$, $\boldsymbol{Y}_2=\left[Y_{12},\ldots,Y_{G2}\right]^T$, $\lambda_{i1}=2N_1m_i\exp\left(\beta_1+\phi_{i1}\right)$, $\lambda_{i2}=2N_2m_i\exp\left(\beta_2+\phi_{i2}\right)$, $C=-\sum_{i=1}^{G}\left[\log\left(Y_{i1}!\right)+\log\left(Y_{i2}!\right)\right]$, and $f\left(\phi_{i1},\phi_{i2}\right)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2\sqrt{1-\rho^2}}\left(\frac{\phi_{i1}^2}{\sigma_1^2}+\frac{\phi_{i2}^2}{\sigma_2^2}-\frac{2\rho\phi_{i1}\phi_{i2}}{\sigma_1\sigma_2}\right)\right]$. We use Monte Carlo integration to evaluate the log-likelihood function. Let $\phi_{i1j}=\sigma_1\xi_{i1j}$ and $\phi_{i2j}=\sigma_2\left(\rho\xi_{i1j}+\sqrt{1-\rho^2}\xi_{i2j}\right)$, where the ξ_{i1j} and ξ_{i2j} are independently and identically distributed random variables following a standard normal distribution. We have

$$l\left(\beta_1,\beta_2,\sigma_1,\sigma_2,\rho|\boldsymbol{Y}_1,\boldsymbol{Y}_2\right)\approx l'\left(\beta_1,\beta_2,\sigma_1,\sigma_2,\rho|\boldsymbol{Y}_1,\boldsymbol{Y}_2\right)=\sum_{i=1}^{G}\log\left[\sum_{j=1}^{M}\exp\left(-\lambda_{i1j}-\lambda_{i2j}\right)\lambda_{i1j}^{Y_{i1}}\lambda_{i2j}^{Y_{i2}}\right]+C,$$

where $\lambda_{i1j}=2N_1m_i\exp\left(\beta_1+\sigma_1\xi_{i1j}\right)$ and $\lambda_{i2j}=2N_2m_i\exp\left[\beta_2+\sigma_2\left(\rho\xi_{i1j}+\sqrt{1-\rho^2}\xi_{i2j}\right)\right]$. Then, we obtain the MLE of $\beta_1,\beta_2,\sigma_1,\sigma_2,\rho$ through maximization of $l'\left(\beta_1,\beta_2,\sigma_1,\sigma_2,\rho|\boldsymbol{Y}_1,\boldsymbol{Y}_2\right)$. Standard error of MLE can be obtained either through inversion of the observed Fisher information matrix or group-wise jackknife if non-invertibility issue occurs.

Computation time

Analysis of a typical pair of disorders with 18,000 genes takes about 10 min on a 2.5 GHz cluster with 1 core.

DNM data and variant annotation

We obtained DNM data from published studies (***Supplementary file 1***-STable 1). DNM data for epileptic encephalopathies from the original release (***Allen et al., 2013***) were not in an editable format and were instead collected from denovo-db (***Turner et al., 2017***). We used ANNOVAR (***Wang et al., 2010***) to annotate all DNMs. Synonymous variants were determined based on the 'synonymous SNV' annotation in ANNOVAR; Variants with 'startloss', 'stopgain', 'stoploss', 'splicing', 'frameshift insertion', 'frameshift deletion', or 'frameshift substitution' annotations were classified as LoF; Dmis variants were defined as nonsynonymous SNVs predicted to be deleterious by MetaSVM ***Dong et al., 2015***; nonsynonymous SNVs predicted to be tolerable by MetaSVM were classified as Tmis. Other DNMs which did not fall into these categories were removed from the analysis. For each variant class, we estimated the mutability of each gene using a sequence-based mutation model (***Samocha et al., 2014***) while adjusting for the sequencing coverage factor based on control trios as previously described (***Jin et al., 2017***; ***Supplementary file 1***-STable 20). We included 18,454 autosomal protein-coding genes in our analysis. *TTN* was removed due to its substantially larger size.

Description and implementation of mTADA

The method mTADA employs a Bayesian framework and estimates the proportion of shared risk genes. Specifically, mTADA assigns all genes into four groups: genes that are not relevant for either disorder, risk genes for the first disorder alone, risk genes for the second disorder alone, and risk genes shared by both disorders. The proportion of these groups are parametrized as π_0,π_1,π_2,π_3, respectively. In particular, parameter π_3 quantifies the extent of genetic sharing between two disorders, with a larger value indicating stronger genetic overlap (***Nguyen et al., 2020***). The 95% credible interval constructed through MCMC is used to measure the uncertainty in π_3 estimates.

The software mTADA requires the following parameters as inputs: proportion of risk genes (π_1^S,π_2^S), mean relative risks ($\bar{\gamma}_1^S,\bar{\gamma}_2^S$), and dispersion parameters ($\bar{\beta}_1^S,\bar{\beta}_2^S$) for both disorders. We used extTADA (***Nguyen et al., 2017***)to estimate these parameters as suggested by the mTADA paper (***Nguyen et al., 2020***). mTADA reported the estimated proportion of shared risk genes π_3 (posterior mode of π_3) and its corresponding 95% credible interval $\left[\mathrm{LB},\ \mathrm{UB}\right]$. We considered $\pi_1^S*\pi_2^S$ as the expected proportion of shared risk genes, and there is significant genetic sharing between two disorders when

$LB > \pi_1^S * \pi_2^S$. We quantify statistical evidence for genetic sharing by comparing the posterior distribution of π_3 with $\pi_1^S * \pi_2^S$,

$$p = 2 * \frac{\sum_{i=1}^{N_{MCMC}} I\left(\pi_3^i < \pi_1^S * \pi_2^S\right)}{N_{MCMC}},$$

where π_3^i is the *i*-th MCMC iteration sample, N_{MCMC} is the number of iterations, and $I()$ is the indicator function. This is also equivalent to performing two-sided inference using posterior probability $P\left(\pi_3 > \pi_1^S * \pi_2^S\right)$. Number of MCMC chain was set as 2 and number of iterations was set as 10,000.

Simulation settings

We assessed the performance of EncoreDNM under the mixed-effects Poisson model. We performed simulations for two variant classes: Tmis and LoF variants, which have the largest and the smallest median mutability values across all genes. First, we performed single-trait simulations to assess estimation precision of elevation parameter β and dispersion parameter σ. We set the true values of β to be −0.5,–0.25, and 0, and the true values of σ to be 0.5, 0.75, and 1. These values were chosen based on the estimated parameters in real DNM data analyses and ensured simulation settings to be realistic. Next, we performed simulations for cross-trait analysis to assess estimation precision of enrichment correlation ρ, whose true values were set to be 0, 0.2, 0.4, 0.6, and 0.8. Sample size for each disorder was set to be 5000. Coverage rate was calculated as the percentage of simulations that the 95% Wald confidence interval covered the true parameter value. Each parameter setting was repeated 100 times.

We also carried out simulations to compare the performance of EncoreDNM and mTADA. False positive rate and statistical power for EncoreDNM were calculated as the proportion of simulation repeats that p-value for enrichment correlation ρ was smaller than 0.05. and the proportion of simulation repeats that p-value for estimated proportion of shared risk genes π_3 was smaller than 0.05 was used for mTADA. We aggregated all variant classes together, so mutability for each gene was determined as the sum of mutabilities across four variant classes (i.e. LoF, Dmis, Tmis, and synonymous).

First, we simulated DNM data under the mixed-effects Poisson model. To see whether two methods would produce false positive findings, we performed simulations under the null hypothesis that the enrichment correlation ρ is zero. We compared two methods under a range of parameter combinations of (β, σ, N) for both disorders: (–0.25, 0.75, 5000) for the baseline setting, (–1, 0.75, 5000) for a setting with small β, (–0.25, 0.5, 5000) for a setting with small σ, and (–0.25, 0.75, 1000) for a setting with small sample size. We also assessed the statistical power of two methods under the alternative hypothesis. True value of enrichment correlation ρ was set to be 0.05, 0.1, 0.15, and 0.2. In the power analysis, parameters (β, σ, N) were fixed at (–0.25, 0.75, 5000) as in the baseline setting when both methods had well-controlled false positive rate.

To ensure a fair comparison, we also compared EncoreDNM and mTADA under a multinomial model, which is different from the data generation processes for the two approaches. For each disorder $(k = 1, 2)$, we randomly selected causal genes of proportion π_k^S . A proportion (i.e. π_3) of causal genes overlap between two disorders. We assumed that the total DNM count to follow a Poisson distribution: $C_k \sim \text{Poisson}\left(u_k * 2N_k \sum_{i=1}^{G} m_i\right)$, where u_k represents an elevation factor to represent systematic bias in the data. Let $\mathbf{Y}_k$ denote the vector of DNMs counts in the exome, $\boldsymbol{m}$ denote the vector of mutability values for all genes, and $\boldsymbol{m}_{causal,\,k}$ denote the vector of mutability with values set to be 0 for non-causal genes of disorder k. We assumed that a proportion p_k of the probands could be attributed to DNMs burden in causal genes, and $1 - p_k$ of the probands obtained DNMs by chance:

$$\mathbf{Y}_k = \mathbf{Y}_{causal,k} + \mathbf{Y}_{background,k},$$
$$\mathbf{Y}_{causal,k} \sim \text{Multinomial}\left(p_k C_k,\ \boldsymbol{m}_{causal,\,k}\right),$$
$$\mathbf{Y}_{background,k} \sim \text{Multinomial}\left(\left(1 - p_k\right) C_k, \boldsymbol{m}\right).$$

To check whether false positive findings could arise, we performed simulations under the null hypothesis that $\pi_3 = \pi_1^S * \pi_2^S$ across a range of parameter combinations of (u, p, N) for both disorders: (0.95, 0.25, 5000) for the baseline setting, (0.75, 0.25, 5000) for a setting with small u (i.e., reduced total mutation count), (0.95, 0.15, 5000) for a setting with small p (fewer probands explained by

DNMs), and (0.95, 0.25, 1000) for a setting with smaller sample size. π_1^S and π_2^S were set as 0.1. We also assessed the statistical power of two methods under the alternative hypothesis that $\pi_3 > \pi_1^S * \pi_2^S$. In power analysis, (u, p, N) were fixed at (0.95, 0.25, 5000) as in the baseline setting when false positive rate for both methods were well-calibrated.

Comparison to the fixed-effects Poisson model

For single-trait analysis, the fixed-effects Poisson model assumes that

$$Y_i \sim \text{Poisson}\left(\lambda_i\right),$$
$$\log\left(\lambda_i\right) = \beta + \log\left(2Nm_i\right), \quad \text{for } i = 1, \ldots, G.$$

Note that the fixed-effects Poisson model is a special case of our proposed mixed-effects Poisson model when $\sigma = 0$. We compared the two models using likelihood ratio test. Under the null hypothesis that $\sigma = 0$, $2\left(l_{alt} - l_{null}\right) \sim \frac{1}{2}\chi_1^2$ asymptotically, where l_{alt} and l_{null} represent the log likelihood of the fitted mixed-effects and fixed-effects Poisson models respectively.

Recurrent genes and DNMs

We used FUMA (***Watanabe et al., 2017***) to perform GO enrichment analysis for genes harboring LoF DNMs in multiple disorders. Due to potential sample overlap between the studies of developmental disorder (***Kaplanis et al., 2020***) and intellectual disability (***Lelieveld et al., 2016***), we excluded intellectual disability from the analysis of recurrent DNMs. We calculated the probability of observing two identical DNMs in two disorders using a Monte Carlo simulation method. For each disorder, we simulated exome-wide DNMs profile from a multinomial distribution, where the size was fixed at the observed DNM count and the per-base mutation probability was determined by the tri-nucleotide base context. We repeated the simulation procedure 100,000 times to evaluate the significance of recurrent DNMs. Lollipop plots for recurrent mutations were generated using MutationMapper on the cBio Cancer Genomics Portal (***Cerami et al., 2012***).

Implementation of cross-trait LD score regression

We used cross-trait LDSC (***Bulik-Sullivan et al., 2015***) to estimate genetic correlations between disorders. LD scores were computed using European samples from the 1000 Genomes Project Phase 3 data (***Auton et al., 2015***). Only HapMap 3 SNPs were used as observations in the explanatory variable with the --merge-alleles flag. Intercepts were not constrained in the analyses.

Estimating enrichment correlation in gene sets

Genes with a high/low probability of intolerance to LoF variants (high-pLI/low-pLI) were defined as the 4,614 genes in the upper/lower quartiles of pLI scores (***Karczewski et al., 2020***). Genes with high/low brain expression (HBE/LBE) were defined as the 4,614 genes in the upper/lower quartiles of expression in the human fetal brain (***Werling et al., 2020***). Genes with high/low heart expression (HHE/LHE) were defined as the 4,614 genes in the upper/lower quartiles of expression in the developing heart of embryonic mouse (***Zaidi et al., 2013***). Five biological pathways have been reported to be involved in congenital heart disease: chromatin remodeling, Notch signaling, cilia function, sarcomere structure and function, and RAS signaling (***Zaidi and Brueckner, 2017***). We extracted 1730 unique genes that belong to these five pathways from the gene ontology database (***Ashburner et al., 2000***) and referred to the union set as CHD-related genes. We repeated EncoreDNM enrichment correlation analysis in these gene sets. One-sided Kolmogorov-Smirnov test was used to assess the statistical difference between enrichment correlation signal strength in different gene sets.

URLs

GWAS summary statistics data of autism spectrum disorder, schizophrenia, and Tourette disorder were downloaded on the PGC website, https://www.med.unc.edu/pgc/download-results/; Summary statistics of cognitive performance were downloaded on the SSGAC website, https://thessgac.com/; Summary statistics of epilepsy were downloaded on the epiGAD website, https://www.epigad.org/; pLI scores were downloaded from gnomAD v3.1 repository https://gnomad.broadinstitute.org/downloads; mTADA, https://github.com/hoangtn/mTADA, ***Nguyen et al.,***

2021; denovo-db, https://denovo-db.gs.washington.edu/denovo-db/; MutationMapper on cBioPortal, https://www.cbioportal.org/mutation_mapper; LDSC, https://github.com/bulik/ldsc; ***Schorsch, 2020***.

Code availability

EncoreDNM software is available at https://github.com/ghm17/EncoreDNM; ***Guo, 2022***.

Acknowledgements

LH acknowledges research support from the National Science Foundation of China (Grant No. 12071243) and Shanghai Municipal Science and Technology Major Project (Grant No. 2017SHZDZX01). QL acknowledges research support from the University of Wisconsin-Madison Office of the Chancellor and the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation and the Waisman Center pilot grant program at University of Wisconsin-Madison. HZ acknowledges research support from the National Institutes of Health (Grant No. R03HD100883 and R01GM134005) and the National Science Foundation (DMS 1902903).

Additional information

Funding

Funder	Grant reference number	Author
National Science Foundation of China	No. 12071243	Lin Hou
Shanghai Municipal Science and Technology Major Project	No. 2017SHZDZX01	Lin Hou
Wisconsin Alumni Research Foundation		Qiongshi Lu
Waisman Center pilot grant program at University of Wisconsin-Madison		Qiongshi Lu
National Institutes of Health	No. R03HD100883 and R01GM134005	Hongyu Zhao
National Science Foundation	DMS 1902903	Hongyu Zhao

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Hanmin Guo, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review and editing; Lin Hou, Qiongshi Lu, Conceptualization, Methodology, Project administration, Supervision, Writing - original draft, Writing - review and editing; Yu Shi, Formal analysis; Sheng Chih Jin, Data curation, Writing - review and editing; Xue Zeng, Boyang Li, Data curation; Richard P Lifton, Martina Brueckner, Validation; Hongyu Zhao, Methodology, Validation, Writing - review and editing

Author ORCIDs

Hanmin Guo http://orcid.org/0000-0001-9022-5307
Qiongshi Lu http://orcid.org/0000-0002-4514-0969

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.75551.sa1
Author response https://doi.org/10.7554/eLife.75551.sa2

Additional files

Supplementary files

- Supplementary file 1. Supplementary Tables 1-20.
- MDAR checklist

Data availability

The current manuscript is a computational study, so no data have been generated for this manuscript.

References

Abdellaoui A, Verweij KJH. 2021. Dissecting polygenic signals from genome-wide association studies on human behaviour. *Nature Human Behaviour* **5**:686–694. DOI: https://doi.org/10.1038/s41562-021-01110-y, PMID: 33986517

Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, Epstein MP, Glauser T, Goldstein DB, Han Y, Heinzen EL, Hitomi Y, Howell KB, Johnson MR, Kuzniecky R, Lowenstein DH, Lu Y-F, Madou MRZ, Marson AG, Mefford HC, et al. 2013. De novo mutations in epileptic encephalopathies. *Nature* **501**:217–221. DOI: https://doi.org/10.1038/nature12439, PMID: 23934111

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**:25–29. DOI: https://doi.org/10.1038/75556

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR, 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526**:68–74. DOI: https://doi.org/10.1038/nature15393, PMID: 26432245

Brainstorm C. 2018. Analysis of shared heritability in common disorders of the brain. *Science (New York, N.Y.)* **360**:aap875. DOI: https://doi.org/10.1126/science.aap875

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, Duncan L, Perry JRB, Patterson N, Robinson EB, Daly MJ, Price AL, Neale BM, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3. 2015. An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**:1236–1241. DOI: https://doi.org/10.1038/ng.3406, PMID: 26414676

Cao M, Wu JI. 2015. Camk2a-Cre-mediated conditional deletion of chromatin remodeler Brg1 causes perinatal hydrocephalus. *Neuroscience Letters* **597**:71–76. DOI: https://doi.org/10.1016/j.neulet.2015.04.041, PMID: 25929186

Cardozo T, Pagano M. 2004. The SCF ubiquitin ligase: insights into a molecular machine. *Nature Reviews. Molecular Cell Biology* **5**:739–751. DOI: https://doi.org/10.1038/nrm1471, PMID: 15340381

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* **2**:401–404. DOI: https://doi.org/10.1158/2159-8290.CD-12-0095, PMID: 22588877

Christensen D, Van Naarden Braun K, Doernberg NS, Maenner MJ, Arneson CL, Durkin MS, Benedict RE, Kirby RS, Wingate MS, Fitzgerald R, Yeargin-Allsopp M. 2014. Prevalence of cerebral palsy, co-occurring autism spectrum disorders, and motor functioning - Autism and Developmental Disabilities Monitoring Network, USA, 2008. *Developmental Medicine and Child Neurology* **56**:59–65. DOI: https://doi.org/10.1111/dmcn.12268, PMID: 24117446

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics* **24**:2125–2137. DOI: https://doi.org/10.1093/hmg/ddu733, PMID: 25552646

Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, Carrera N, Humphreys I, Johnson JS, Roussos P, Barker DD, Banks E, Milanova V, Grant SG, Hannon E, Rose SA, et al. 2014. De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**:179–184. DOI: https://doi.org/10.1038/nature12929, PMID: 24463507

Garne E, Dolk H, Krägeloh-Mann I, Holst Ravn S, Cans C, SCPE Collaborative Group. 2008. Cerebral palsy and congenital malformations. *European Journal of Paediatric Neurology* **12**:82–88. DOI: https://doi.org/10.1016/j.ejpn.2007.07.001, PMID: 17881257

Gratten J, Wray NR, Keller MC, Visscher PM. 2014. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience* **17**:782–790. DOI: https://doi.org/10.1038/nn.3708, PMID: 24866044

Gregor A, Sadleir LG, Asadollahi R, Azzarello-Burri S, Battaglia A, Ousager LB, Boonsawat P, Bruel A-L, Buchert R, Calpena E, Cogné B, Dallapiccola B, Distelmaier F, Elmslie F, Faivre L, Haack TB, Harrison V, Henderson A, Hunt D, Isidor B, et al. 2018. De Novo Variants in the F-Box Protein FBXO11 in 20 Individuals with a Variable Neurodevelopmental Disorder. *American Journal of Human Genetics* **103**:305–316. DOI: https://doi.org/10.1016/j.ajhg.2018.07.003, PMID: 30057029

Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, Ip HF, Marioni RE, McIntosh AM, Deary IJ, Koellinger PD, Harden KP, Nivard MG, Tucker-Drob EM. 2019. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour* **3**:513–525. DOI: https://doi.org/10.1038/s41562-019-0566-x, PMID: 30962613

Grotzinger AD, Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewis CM, McIntosh AM, Grove J, Dalsgaard S, Lesch KP, Strom N, Meier SM, Mattheisen M, Børglum AD, Mors O, Breen G, Lee PH, Kendler KS, Smoller JW, Tucker-Drob EM, et al. 2020. Genetic Architecture of 11 Major Psychiatric Disorders at Biobehavioral, Functional Genomic, and Molecular Genetic Levels of Analysis. medRxiv. DOI: https://doi.org/10.1101/2020.09.22.20196089

Gulisano M, Calì PV, Cavanna AE, Eddy C, Rickards H, Rizzo R. 2011. Cardiovascular safety of aripiprazole and pimozide in young patients with Tourette syndrome. *Neurological Sciences* **32**:1213–1217. DOI: https://doi.org/10.1007/s10072-011-0678-1, PMID: 21732066

Guo H, Li JJ, Lu Q, Hou L. 2021. Detecting local genetic correlations with scan statistics. *Nature Communications* **12**:2033. DOI: https://doi.org/10.1038/s41467-021-22334-6, PMID: 33795679

Guo H. 2022. EncoreDNM. swh:1:rev:44ec5903b4c34e7b73ed7791f30d0b3544bafcd1. GitHub. https://github.com/ghm17/EncoreDNM

Hoischen A, Krumm N, Eichler EE. 2014. Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nature Neuroscience* **17**:764–772. DOI: https://doi.org/10.1038/nn.3703, PMID: 24866042

Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, DePalma SR, McKean D, Wakimoto H, Gorham J, Jin SC, Deanfield J, Giardini A, Porter GA Jr, Kim R, Bilguvar K, López-Giráldez F, Tikhonova I, Mane S, Romano-Adesman A, et al. 2015. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science (New York, N.Y.)* **350**:1262–1266. DOI: https://doi.org/10.1126/science.aac9396, PMID: 26785492

Hormozdiari F, Penn O, Borenstein E, Eichler EE. 2015. The discovery of integrated gene networks for autism and related disorders. *Genome Research* **25**:142–154. DOI: https://doi.org/10.1101/gr.178855.114, PMID: 25378250

Howrigan DP, Rose SA, Samocha KE, Fromer M, Cerrato F, Chen WJ, Churchhouse C, Chambert K, Chandler SD, Daly MJ, Dumont A, Genovese G, Hwu H-G, Laird N, Kosmicki JA, Moran JL, Roe C, Singh T, Wang S-H, Faraone SV, et al. 2020. Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nature Neuroscience* **23**:185–193. DOI: https://doi.org/10.1038/s41593-019-0564-3, PMID: 31932770

Jansen S, van der Werf IM, Innes AM, Afenjar A, Agrawal PB, Anderson IJ, Atwal PS, van Binsbergen E, van den Boogaard M-J, Castiglia L, Coban-Akdemir ZH, van Dijck A, Doummar D, van Eerde AM, van Essen AJ, van Gassen KL, Guillen Sacoto MJ, van Haelst MM, Iossifov I, Jackson JL, et al. 2019. De novo variants in FBXO11 cause a syndromic form of intellectual disability with behavioral problems and dysmorphisms. *European Journal of Human Genetics* **27**:738–746. DOI: https://doi.org/10.1038/s41431-018-0292-2, PMID: 30679813

Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, Zeng X, Qi H, Chang W, Sierant MC, Hung W-C, Haider S, Zhang J, Knight J, Bjornson RD, Castaldi C, Tikhonoa IR, Bilguvar K, Mane SM, Sanders SJ, et al. 2017. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nature Genetics* **49**:1593–1601. DOI: https://doi.org/10.1038/ng.3970, PMID: 28991257

Jin SC, Dong W, Kundishora AJ, Panchagnula S, Moreno-De-Luca A, Furey CG, Allocco AA, Walker RL, Nelson-Williams C, Smith H, Dunbar A, Conine S, Lu Q, Zeng X, Sierant MC, Knight JR, Sullivan W, Duy PQ, DeSpenza T, Reeves BC, et al. 2020a. Exome sequencing implicates genetic disruption of prenatal neuro-gliogenesis in sporadic congenital hydrocephalus. *Nature Medicine* **26**:1754–1765. DOI: https://doi.org/10.1038/s41591-020-1090-2, PMID: 33077954

Jin SC, Lewis SA, Bakhtiari S, Zeng X, Sierant MC, Shetty S, Nordlie SM, Elie A, Corbett MA, Norton BY, van Eyk CL, Haider S, Guida BS, Magee H, Liu J, Pastore S, Vincent JB, Brunstrom-Hernandez J, Papavasileiou A, Fahey MC, et al. 2020b. Mutations disrupting neuritogenesis genes confer risk for cerebral palsy. *Nature Genetics* **52**:1046–1056. DOI: https://doi.org/10.1038/s41588-020-0695-1, PMID: 32989326

Kaplanis J, Samocha KE, Wiel L, Zhang Z, Arvai KJ, Eberhardt RY, Gallone G, Lelieveld SH, Martin HC, McRae JF, Short PJ, Torene RI, de Boer E, Danecek P, Gardner EJ, Huang N, Lord J, Martincorena I, Pfundt R, Reijnders MRF, et al. 2020. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**:757–762. DOI: https://doi.org/10.1038/s41586-020-2832-5, PMID: 33057194

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**:434–443. DOI: https://doi.org/10.1038/s41586-020-2308-7, PMID: 32461654

Kielinen M, Rantala H, Timonen E, Linna SL, Moilanen I. 2004. Associated medical disorders and disabilities in children with autistic disorder: a population-based study. *Autism* **8**:49–60. DOI: https://doi.org/10.1177/1362361304040638, PMID: 15070547

Kilincaslan A, Mukaddes NM. 2009. Pervasive developmental disorders in individuals with cerebral palsy. *Developmental Medicine and Child Neurology* **51**:289–294. DOI: https://doi.org/10.1111/j.1469-8749.2008.03171.x, PMID: 19335564

Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, Raja A, Coe BP, Stessman HA, He Z-X, Leal SM, Bernier R, Eichler EE. 2015. Excess of rare, inherited truncating mutations in autism. *Nature Genetics* **47**:582–588. DOI: https://doi.org/10.1038/ng.3303, PMID: 25961944

Kume T, Deng KY, Winfrey V, Gould DB, Walter MA, Hogan BL. 1998. The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. *Cell* **93**:985–996. DOI: https://doi.org/10.1016/s0092-8674(00)81204-0, PMID: 9635428

Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. 2012. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics (Oxford, England)* **28**:2540–2542. DOI: https://doi.org/10.1093/bioinformatics/bts474, PMID: 22843982

Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, Absher D, Agartz I, Akil H, Amin F, Andreassen OA, Anjorin A, Anney R, Anttila V, Arking DE, Asherson P, et al. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* **45**:984–994. DOI: https://doi.org/10.1038/ng.2711, PMID: 23933821

Lee PH. 2019. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**:1469–1482. DOI: https://doi.org/10.1016/j.cell.2019.11.020, PMID: 31380891

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**:285–291. DOI: https://doi.org/10.1038/nature19057, PMID: 27535533

Lelieveld SH, Reijnders MRF, Pfundt R, Yntema HG, Kamsteeg E-J, de Vries P, de Vries BBA, Willemsen MH, Kleefstra T, Löhner K, Vreeburg M, Stevens SJC, van der Burgt I, Bongers EMHF, Stegmann APA, Rump P, Rinne T, Nelen MR, Veltman JA, Vissers LELM, et al. 2016. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nature Neuroscience* **19**:1194–1196. DOI: https://doi.org/10.1038/nn.4352, PMID: 27479843

Li J, Cai T, Jiang Y, Chen H, He X, Chen C, Li X, Shao Q, Ran X, Li Z, Xia K, Liu C, Sun ZS, Wu J. 2016. Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Molecular Psychiatry* **21**:290–297. DOI: https://doi.org/10.1038/mp.2015.40, PMID: 25849321

Lu Q, Li B, Ou D, Erlendsdottir M, Powles RL, Jiang T, Hu Y, Chang D, Jin C, Dai W, He Q, Liu Z, Mukherjee S, Crane PK, Zhao H. 2017. A Powerful Approach to Estimating Annotation-Stratified Genetic Covariance via GWAS Summary Statistics. *American Journal of Human Genetics* **101**:939–964. DOI: https://doi.org/10.1016/j.ajhg.2017.11.001, PMID: 29220677

Lumenta CB, Skotarczak U. 1995. Long-term follow-up in 233 patients with congenital hydrocephalus. *Child's Nervous System* **11**:173–175. DOI: https://doi.org/10.1007/BF00570260, PMID: 7773979

Munkin MK, Trivedi PK. 1999. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. *The Econometrics Journal* **2**:29–48. DOI: https://doi.org/10.1111/1368-423X.00019

Nguyen HT, Bryois J, Kim A, Dobbyn A, Huckins LM, Munoz-Manchado AB, Ruderfer DM, Genovese G, Fromer M, Xu X, Pinto D, Linnarsson S, Verhage M, Smit AB, Hjerling-Leffler J, Buxbaum JD, Hultman C, Sklar P, Purcell SM, Lage K, et al. 2017. Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *Genome Medicine* **9**:114. DOI: https://doi.org/10.1186/s13073-017-0497-y, PMID: 29262854

Nguyen T-H, Dobbyn A, Brown RC, Riley BP, Buxbaum JD, Pinto D, Purcell SM, Sullivan PF, He X, Stahl EA. 2020. mTADA is a framework for identifying risk genes from de novo mutations in multiple traits. *Nature Communications* **11**:2929. DOI: https://doi.org/10.1038/s41467-020-16487-z, PMID: 32522981

Nguyen TH, Dobbyn A, Brown RC, Riley BP, Buxbaum J, Pinto D, Purcell SM, Sullivan PF, He X, Eli A. 2021. mTADA is a framework for identifying risk genes from de novo mutations in multiple traits. 7630c4b. GitHub. https://github.com/hoangtn/mTADA

Ning Z, Pawitan Y, Shen X. 2020. High-definition likelihood inference of genetic correlations across human complex traits. *Nature Genetics* **52**:859–864. DOI: https://doi.org/10.1038/s41588-020-0653-y, PMID: 32601477

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, et al. 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**:246–250. DOI: https://doi.org/10.1038/nature10989, PMID: 22495309

Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. 2016. Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics* **48**:709–717. DOI: https://doi.org/10.1038/ng.3570, PMID: 27182965

Rees E, Creeth HDJ, Hwu H-G, Chen WJ, Tsuang M, Glatt SJ, Rey R, Kirov G, Walters JTR, Holmans P, Owen MJ, O'Donovan MC. 2021. Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations. *Nature Communications* **12**:5353. DOI: https://doi.org/10.1038/s41467-021-25532-4, PMID: 34504065

Reid SM, Meehan EM, Arnup SJ, Reddihough DS. 2018. Intellectual disability in cerebral palsy: a population-based retrospective study. *Developmental Medicine and Child Neurology* **60**:687–694. DOI: https://doi.org/10.1111/dmcn.13773, PMID: 29667705

Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnström K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM, Palotie A, Boerwinkle E, Buxbaum JD, Cook EH Jr, Gibbs RA, et al. 2014. A framework for the interpretation of de novo mutation in human disease. *Nature Genetics* **46**:944–950. DOI: https://doi.org/10.1038/ng.3050, PMID: 25086666

Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, Peng M, Collins R, Grove J, Klei L, Stevens C, Reichert J, Mulhern MS, Artomov M, Gerges S, Sheppard B, Xu X, Bhaduri A, Norman U, Brand H, et al. 2020. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**:568-584.. DOI: https://doi.org/10.1016/j.cell.2019.12.036, PMID: 31981491

Schorsch E. 2020. LDSC (LD SCore) v1.0.1. aa33296. GitHub. https://github.com/bulik/ldsc

Shi H, Mancuso N, Spendlove S, Pasaniuc B. 2017. Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *American Journal of Human Genetics* **101**:737–751. DOI: https://doi.org/10.1016/j.ajhg.2017.09.022, PMID: 29100087

Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, Wedow R, Zacher M, Furlotte NA, 23andMe Research Team, Social Science Genetic Association Consortium, Magnusson P, Oskarsson S, Johannesson M, Visscher PM, Laibson D, Cesarini D, Neale BM, Benjamin DJ. 2018. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics* **50**:229–237. DOI: https://doi.org/10.1038/s41588-017-0009-4, PMID: 29292387

Turner TN, Yi Q, Krumm N, Huddleston J, Hoekzema K, F Stessman HA, Doebley A-L, Bernier RA, Nickerson DA, Eichler EE. 2017. denovo-db: A compendium of human de novo variants. *Nucleic Acids Research* **45**:D804–D811. DOI: https://doi.org/10.1093/nar/gkw865, PMID: 27907889

Veltman JA, Brunner HG. 2012. De novo mutations in human genetic disease. *Nature Reviews. Genetics* **13**:565–575. DOI: https://doi.org/10.1038/nrg3241, PMID: 22805709

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**:e164. DOI: https://doi.org/10.1093/nar/gkq603, PMID: 20601685

Wang Q, Yang C, Gelernter J, Zhao H. 2015. Pervasive pleiotropy between psychiatric disorders and immune disorders revealed by integrative analysis of multiple GWAS. *Human Genetics* **134**:1195–1209. DOI: https://doi.org/10.1007/s00439-015-1596-8, PMID: 26340901

Watanabe K, Taskesen E, van Bochoven A, Posthuma D. 2017. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**:1–11. DOI: https://doi.org/10.1038/s41467-017-01261-5, PMID: 29184056

Wei Q, Zhan X, Zhong X, Liu Y, Han Y, Chen W, Li B. 2015. A Bayesian framework for de novo mutation calling in parents-offspring trios. *Bioinformatics (Oxford, England)* **31**:1375–1381. DOI: https://doi.org/10.1093/bioinformatics/btu839, PMID: 25535243

Werling DM, Pochareddy S, Choi J, An JY, Sheppard B, Peng M, Li Z, Dastmalchi C, Santpere G, Sousa AMM, Tebbenkamp ATN, Kaur N, Gulden FO, Breen MS, Liang L, Gilson MC, Zhao X, Dong S, Klei L, Cicek AE, et al. 2020. Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic Coordination in the Developing Human Prefrontal Cortex. *Cell Reports* **31**:e107489. DOI: https://doi.org/10.1016/j.celrep.2020.03.053, PMID: 32268104

Willsey AJ, Fernandez TV, Yu D, King RA, Dietrich A, Xing J, Sanders SJ, Mandell JD, Huang AY, Richer P, Smith L, Dong S, Samocha KE, Tourette International Collaborative Genetics (TIC Genetics), Tourette Syndrome Association International Consortium for Genetics (TSAICG), Neale BM, Coppola G, Mathews CA, Tischfield JA, Scharf JM, et al. 2017. De Novo Coding Variants Are Strongly Associated with Tourette Disorder. *Neuron* **94**:486-499.. DOI: https://doi.org/10.1016/j.neuron.2017.04.024, PMID: 28472652

Willsey AJ, Morris MT, Wang S, Willsey HR, Sun N, Teerikorpi N, Baum TB, Cagney G, Bender KJ, Desai TA, Srivastava D, Davis GW, Doudna J, Chang E, Sohal V, Lowenstein DH, Li H, Agard D, Keiser MJ, Shoichet B, et al. 2018. The Psychiatric Cell Map Initiative: A Convergent Systems Biological Approach to Illuminating Key Molecular Pathways in Neuropsychiatric Disorders. *Cell* **174**:505–520. DOI: https://doi.org/10.1016/j.cell.2018.06.016, PMID: 30053424

Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, Romano-Adesman A, Bjornson RD, Breitbart RE, Brown KK, Carriero NJ, Cheung YH, Deanfield J, DePalma S, Fakhro KA, Glessner J, Hakonarson H, Italia MJ, Kaltman JR, Kaski J, et al. 2013. De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**:220–223. DOI: https://doi.org/10.1038/nature12141, PMID: 23665959

Zaidi S, Brueckner M. 2017. Genetics and Genomics of Congenital Heart Disease. *Circulation Research* **120**:923–940. DOI: https://doi.org/10.1161/CIRCRESAHA.116.309140, PMID: 28302740

Zhang Y, Cheng Y, Jiang W, Ye Y, Lu Q, Zhao H. 2021a. Comparison of methods for estimating genetic correlation between complex traits using GWAS summary statistics. *Briefings in Bioinformatics* **22**:bbaa442. DOI: https://doi.org/10.1093/bib/bbaa442

Zhang Y, Lu Q, Ye Y, Huang K, Liu W, Wu Y, Zhong X, Li B, Yu Z, Travers BG, Werling DM, Li JJ, Zhao H. 2021b. SUPERGNOVA: local genetic correlation analysis reveals heterogeneous etiologic sharing of complex traits. *Genome Biology* **22**:1–30. DOI: https://doi.org/10.1186/s13059-021-02478-w, PMID: 34493297