BrainXcan identifies brain features associated with behavioral and psychiatric traits using large scale genetic and imaging data

Yanyu Liang[1,*] Owen Melia[2] Thomas Brettin[3,4] Andrew Brown[5]
Hae Kyung Im[1,3,*]

1 Section of Genetic Medicine, University of Chicago, Chicago, Illinois, United States of America
2 Department of Computer Science, University of Chicago, Chicago, Illinois, United States of America
3 Computing Environment and Life Sciences Directorate, Argonne National Laboratory, Argonne, Illinois, United States of America
4 Consortium for Advanced Science and Engineering, University of Chicago, Chicago, Illinois, United States of America
5 Department of Population Health and Genomics, University of Dundee, Dundee, United Kingdom
∗ Correspondence to yanyul@uchicago.edu and haky@uchicago.edu

Abstract

Advances in brain MRI have enabled many discoveries in neuroscience. Comparison of brain MRI features between cases and controls have highlighted potential causes of psychiatric and behavioral traits. However, due to the cost of collecting MRI data and the difficulty in recruiting particular patient groups, most studies have small sample sizes, limiting their reliability. Furthermore, interpretation is complicated by reverse causality, where many observed brain differences are the result of disease rather than the cause. Here we propose a method (BrainXcan) that leverages the power of large-scale genome-wide association studies (GWAS), reference brain MRI data, and methodological advances in causal inference using genetic instruments to discover new mechanisms of disease etiology and validate existing ones. BrainXcan tests complex traits for association with genetic predictors of brain MRI derived phenotypes to pinpoint relevant region-specific and cross-brain features. It also evaluates consistency and direction of the causal flow with Mendelian Randomization. As this approach requires only genetic data, BrainXcan allows us to test a host of hypotheses on mental illness, across many disorders and MRI modalities, using existing public data resources. Our method shows that reduced axonal density across the brain is associated with the risk of schizophrenia, consistent with the disconnectivity hypothesis. We also find structural features hippocampus, amygdala, and anterior cingulate cortex among others associated with schizophrenia risk highlighting the potential of our approach to bring orthogonal lines of evidence to inform the biology of

complex traits.

Introduction

Advances in brain MRI have enabled many discoveries in neuroscience. However, reproducibility of brain-wide associations studies (BWAS) is low due, in large part, to small sample sizes (Marek et al., 2020). These small sample sizes are the result of the high cost of collecting MRI scans, as well as the difficulty in recruiting patients with particular mental illnesses. Also, unlike genome-wide association studies where disease status does not alter germline genetic variation, brain features can be altered by disease status and treatments, which can yield significant associations due to reverse causality.

The UK Biobank is in the process of measuring brain MRI in 100,000 individuals (Littlejohns et al., 2020). The unprecedented scale of the data, the automated uniform processing of the data, the availability of genetic and a myriad of phenotypic data will undoubtedly catalyze many discoveries in the coming years. The interim analysis of brain MRI image derived phenotypes (IDPs) found many genome-wide significant loci associated and established that most IDPs are heritable (Smith et al., 2021). Zhao et al generated polygenic risk scores of 101 brain volumetric phenotypes using 19,629 UK Biobank participant data and showed they could explain more than 6% of the phenotypic variance in four independent studies (Zhao et al., 2019).

The Psychiatric Genomics Consortium is a cooperative effort of investigators across the world that combines studies of many mental disorders and has enabled discoveries that would not have been possible within each of the studies. All their GWAS summary results are publicly available to allow other investigators to test their own hypotheses and extract new biological insight. The PGC studies 11 psychiatric disorders including ADHD, Alzheimer's disease, autism, bipolar disorder, and schizophrenia.

Methods that leverage UK Biobank's large scale image data and the PGC's large scale GWAS data have the potential to unlock many insights into the biology of mental disorders. In this paper we propose one such method, BrainXcan, which leverages these two data resources to address some of the deficiencies in small scale MRI studies. Using UK Biobank data as a reference, we build models to predict brain IDPs from genetic data. These models can then be applied to from genome-wide association studies. For example, using the schizophrenia GWAS data collected by the PGC, our method tests for association between schizophrenia and a number of different functional, structural and diffusion MR modalities with size of $\sim 70,000$ cases and $\sim 240,000$ controls. Furthermore, by applying a Mendelian randomization approach we infer the direction of causality: whether the changes in IDP are the cause of disease or a consequence of it.

IDP-associated genetic markers have been used for causal inference with methods such as Mendelian Ran-

domization to investigate the mediating role of brain features on behavioral phenotypes with both large sample sizes and protection from reverse causality. For instance, Jansen et al. (2020) studied the genomic loci and corresponding genes that are shared between brain volume IDPs and intelligence and they identified 92 shared genes which provided insight of the shared genetic etiology of brain volume and intelligence. Shen et al. (2020) performed bi-directional MR analysis with depression and dMRI IDPs finding suggestive evidence that the change of the mean diffusivity in thalamic radiations could be a consequence of major depressive disorder. A related approach is one that correlates genetically predicted brain IDP/phenotype and the complex trait, an extension of transcriptome-based methods (Gamazon et al., 2015; Gusev et al., 2016) to IDPs. Based on this idea, Knutson et al. (2020) developed imaging-wide association study (IWAS) using 14 brain features from the Alzheimer's Disease Neuroimaging Initiative. They also used standard PRS approaches to generate prediction weights using the GWAS summary results from Elliott et al. (2018) (n=8,428).

In this paper, we perform an in-depth analysis of the genetic architecture of IDPs and further process UK Biobank's IDPs to develop a framework that maximizes interpretability, robustness, computational efficiency, and user friendliness.

The high polygenicity of brain features imposes several challenges to existing methods limiting the power to detect their link to diseases; strong genetic instruments needed for Mendelian randomization based approaches are difficult to identify. We address these challenges by developing polygenic predictors of IDPs informed by their complex genetic architecture and correlation structure. To facilitate interpretation of the results, we develop region-specific and brain-wide predictors providing an in-depth analysis and quantification of potential biases. We make sure that the implementation is computationally efficient and scalable to genome-wide Biobank-scale data. We develop an extension of the association method that can infer the association using the increasingly available GWAS summary results, i.e. without the need to use individual level data. We add a Mendelian Randomization module to estimate the direction of the causal flow. We illustrate the power of the approach by applying it to 44 human traits. Finally, we provide the software, the recommended pipeline, and automated reports to improve usability and lower the barrier to adoption for users less familiar with genetic studies.

Results

Overview of the BrainXcan framework

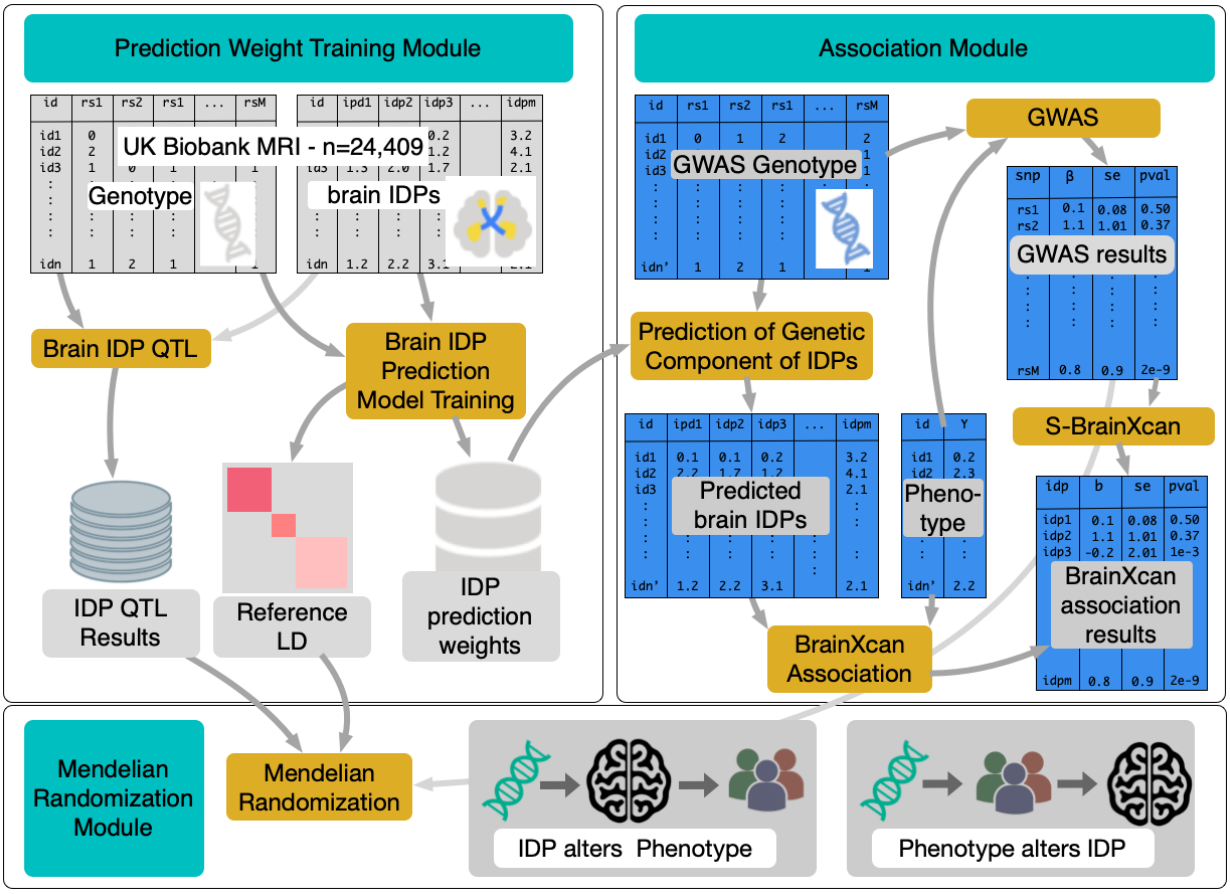


Figure 1: The workflow for implementing BrainXcan framework. The workflow is divided into three modules: "Prediction weight training", "Association between predicted IDP and phenotype", and "Mendelian randomization". The "Prediction weight training" module uses reference brain MRI data with genotype and IDPs and trains genetic predictors of brain IDPs. It also computes QTLs of brain IDPs and LD information, i.e. genotype covariance. This module only needs to be run once and the results are provided to users as part of the BrainXcan software package. The "Association" and "Mendelian randomization" modules are the main analysis components that will be performed by most BrainXcan users. In a typical use case, users first run "Association" module with either individual-level phenotype/genotype data or the GWAS summary statistics. The significant (with user-defined threshold) IDPs, will be processed with the "Mendelian randomization" module to examine the direction of the putative causal flow and the consistency across multiple loci.

The BrainXcan framework is organized in three modules as outlined in Fig. 1. The **"Prediction weight training" module** trains linear genetic predictors of brain IDPs, performs brain IDP QTL analysis (association between brain IDPs and genotype) and calculates the sample covariance of the genotypes. These outcomes are saved for use in subsequent modules and shared publicly in `predictdb.org` with versioned and permanent record in zenodo.org.

The **"Association" module** operates on the GWAS of the phenotype of interest where genotype and phenotype data are available. First, it computes the genetic predictors of brain IDPs using the genotype data and the weights from the training stage. Next, it calculates the association statistics between the predicted brain IDPs and the phenotype of interest via linear regression. Generlized linear models to take into account binary or other types of outcomes can be easily accommodated. Significant associations pinpoint candidate causal relationship between brain features and the phenotype. As explained in the IDP processing section, we use further derived IDPs that represent region-specific features as well as brain-wide features.

For most large scale GWAS studies, the individual level data are not available but the BrainXcan framework can still be applied because the association statistics can be inferred using the summary results from GWAS, the IDP model weights, and the reference LD data generated by the first module (see Methods). The feasibility of this approach was shown with the PrediXcan framework (Barbeira et al., 2018).

The **"Mendelian randomization" module** performs a number of multiple instrument-based Mendelian randomizations to determine the direction of the putative causal flow, i.e. whether alteration in brain features affects the complex trait or whether the phenotype status alters brain features. It provides bi-directional test of causal flow and effect size scatter plots to help assess the consistency of the results.

Preprocessing brain MRI derived phenotypes

We downloaded uniformly preprocessed brain MRI derived phenotypes from the UK Biobank (IDPs). We focused here on 159 IDPs derived from structural images representing total and gray matter volumes of different regions of the brain and 300 diffusion MRI derived phenotypes representing neurite density, dispersion, and connectivity features. After excluding related individuals and those of non European ancestry, IDP data on 24,409 individuals remained. We adjusted for covariates including the first 10 genetic PCs, age at recruitment, sex, and four technical covariates indicating the location of the head in the scanner. See details in (Methods) and the list of IDPs in table S1.

We processed the structural and diffusion MRI modalities separately. We further categorized the structural IDPs into 5 subtypes: cortical gray matter volume, sub-cortical gray matter volume, sub-cortical total volume, cerebellum gray matter, and brain-stem. We also included total gray matter volume and total gray and white matter volume as additional IDPs. Diffusion MRI measures were also categorized into 4 subtypes:

fractional anisotropy (FA, a measure of diffusion along the white matter tracts), intracellular volume fraction (ICVF, an estimate of neurite/axonal density), isotropic volume fraction (ISOVF, an index of the relative extra-cellular water diffusion), and orientation diffusion index (OD, a measure of neurite dispersion) (Zhang et al., 2012). Our main analysis focused on the Tract-based spatial statistics (TBSS) IDPs (192 in total) while the probability track based measures (108 in total) were left for internal validation.

For each subtype, we performed a principal components analysis. The first principal component was a weighted average of IDPs in the subtype suggesting that they could be used as proxies for the brain-wide feature (fig. S1). For diffusion MRI subtypes, the principal component explained 39% (FA), 53% (ICVF), 26% (ISOVF), and 20% (OD) of the total variance. For the structural MRIs, the first principal component explained 16% (cortical gray matter), 34% (sub-cortical gray matter), 37% (sub-cortical total volume), and 45% (cerebellum gray matter) of the total variability. The residual IDPs after adjusting for the first principal component within each subtype, had a much reduced correlation structure as shown in fig. S2 and S3.

Brain IDPs can be decomposed into common and region-specific features

Since an important attribute of any method is the interpretability of the results, we sought to define brain features with the goal of facilitating interpretation. We prioritized the ability to tease out brain-wide effects from region-specific effects. i.e., determining whether a detected association with the phenotype was due to a feature that is common across the whole brain or specific to a region (e.g. thalamus, anterior limb of the internal capsule, etc.).

To distinguish between brain-wide effects and region-specific effects, we postulated the generative model shown in Fig. 2a. The brain feature in each region (F_k) was modeled as the sum of two independent latent components: one region-specific (R_k) and one brain-wide (L). The observed value, IDP, was modeled as a noisy version of the region's feature ($\text{IDP}_k = F_k + \epsilon_k = L + R_k + \epsilon_k$). The parameter s^2 determined the scale of the region-specific component (modeled as a normal random variable with variance s^2) and t^2 was the variance of noise term ϵ_k.

Attenuation and collider biases can be estimated

Our effects of interest were represented as β_k (region-specific) and α (brain-wide). Ideally, we would like to fit a regression model of the trait Y jointly on the brain-wide (L) and the region-specific (R_k) components but since the latent variables were not available to us, we used the principal component (PC) and the residual IDPs (residual of IDP_k after regressing out the subtype's PC) as proxies.

To examine the effect of using these proxies instead of the latent factors R_k and L, we derived analytical expressions for the bias when regressing the trait on one of the residual IDPs ($Y \sim \text{resIDP}_k$) and the

subytpe's PC ($Y \sim \text{PC}$) (see details in Supplementary Notes 1).

The overall noise level within IDP_k led to what is known as attenuation bias whereas adjusting for PCs (weighted sum of IDPs) led to what is known as collider bias (Dahl et al., 2019).

$$\text{coef IDP}_k = \beta_k - \frac{t^2}{s^2+t^2} \cdot \beta_k - \frac{s^2}{s^2+t^2} \sum_{j \neq k} \frac{\beta_j}{m-1} + O_p\left(\frac{1}{n}\right), \tag{1}$$

where m is the number of IDPs in the subtype (ranging from 14 to 96) and n is the sample size of the GWAS study or equivalently of the BrainXcan association study (typically $\sim 100K - 1M$). The attenuation bias reduced the estimate of β_k by a factor of $(1 - \frac{t^2}{s^2+t^2})$. The third term represents the collider bias, proportional to the average effect size across brain regions, which can be expected to be relatively small if the effect is specific to a region.

Regression on the subtype's PC instead of the latent variable L yields a biased coefficient:

$$\text{coef PC} = \alpha + \sum_j \beta_j + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{1}{n}\right). \tag{2}$$

This coefficient is the sum of the latent brain-wide effect (α) and the effects of individual regions. The interpretation of this coefficient will depend on the significance of the individual region effects estimated in Eq.(1). If all region-specific effects are small and not significant, then we can assume that the brain-wide effect α is not biased. However, if the attenuation bias is very large, it is possible that the effect is missed in the first regression (Y on residaul IDP) but detected in Eq.(2). These considerations must be carefully taken into account for the interpretation.

Next, to better inform optimal prediction approaches, we proceeded to investigate the genetic architecture of these brain features by calculating their heritability and degree of polygenicity.

Both global and region-specific brain features are heritable and highly polygenic

We calculated the heritability of brain IDPs using standard mixed effects modeling approach (Yang et al., 2010) (Methods). Heritability estimates ranged from 5% to 43% with all the 95% confidence intervals above zero as shown in Fig. 3a. Since principal components were heritable, the residual IDPs (PC-adjusted) were less heritable than the original IDPs. (fig. S4).

To quantify the degree of polygenicity of brain IDPs, we estimated the effective number of independently associated SNPs (M_e) using the stratified LD fourth moments regression (O'Connor et al., 2019) (Methods and fig. S5). Two hundred and sixty five out of 359 IDPs yielded a significant ($p < 0.05$) estimate of M_e with values ranging from 1,035 to 24,675 with a median of 6,245 which was higher than the estimates

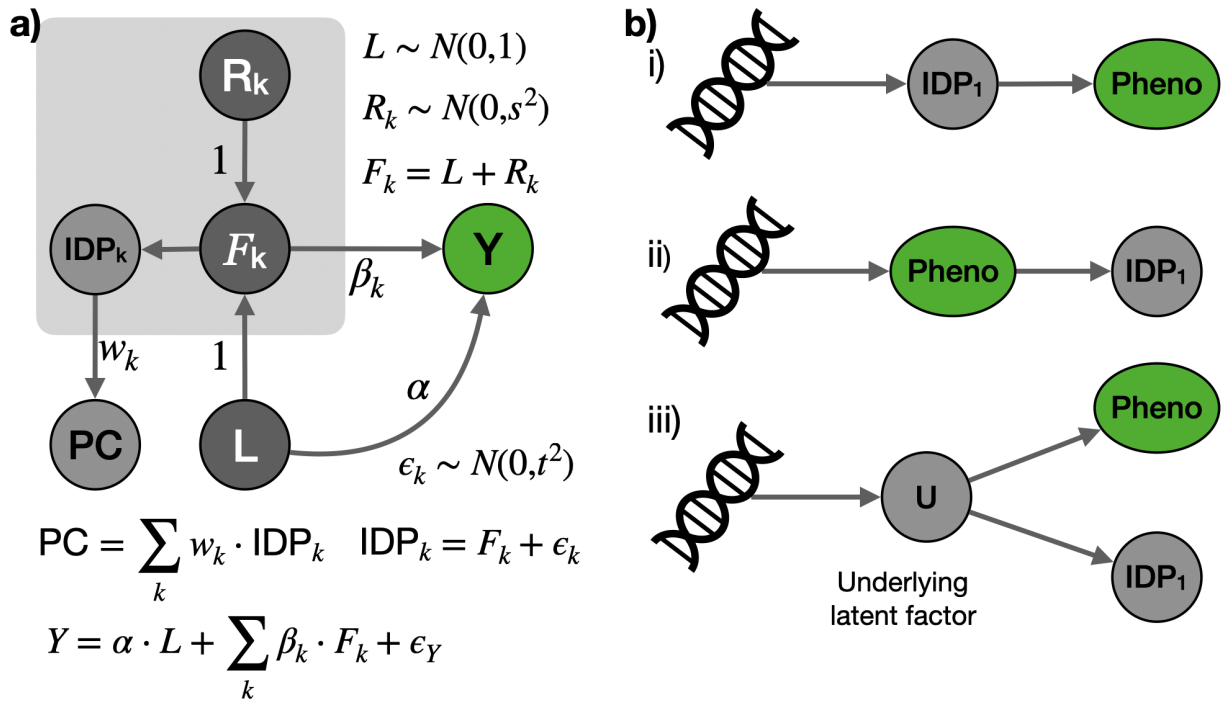


Figure 2: Generative model of brain features and complex traits. a) Brain features, F_k, were modeled as a sum of a region-specific effect R_k and brain-wide component L. The IDPs were considered to be noisy measurements of brain features. Principal components of IDPs were weighted averages of IDPs. The complex trait Y was modeled as the sum of brain-wide effect (α) and region-specific effects (β_k) and an error term (ϵ_Y). **b) Mendelizan Randomization causal flow interpretation.** Associations between brain IDPs and the phenotype can arise from multiple mediating scenarios. We considered i) brain IDP alters phenotype, ii) phenotype (disease status) alters brain IDP, iii) underlying latent factor alters both phenotype and brain IDP. Given the power differential with current GWAS and reference image datasets, significant scenario ii) may not rule out scenario i) or iii). See discussion in text.

for canonical examples of polygenic traits such as height or BMI. The estimates are shown in Fig. 3b with common human traits added for reference.

Ridge regression predicts brain features better than elastic net

We trained genetic predictors of the original brain IDPs, the derived principal components, and the residual IDPs using penalized regression approaches, ridge regression and elastic net. We restricted our search to linear models so that BrainXcan could be applied using solely GWAS summary statistics even when the individual level genotype and phenotype data were not available. Given the high polygenicity of IDPs, we anticipated that many of the predictors would be based on ridge regression with all the SNPs having a nonzero weight. Therefore, to keep the computation manageable and to make the prediction models applicable to a broader set of GWAS studies where access to individual level data may not be possible, we restricted the training to HapMap3 SNPs, which tend to be imputed with high quality, with MAF > 0.01 in European samples. To avoid issues with strand mis-specification, we excluded the ambiguous SNPs (e.g. AT, CG). These restrictions left us with a total of 1.07 million SNPs (Methods) for the subsequent procedure.

We trained two sets of models, one with a ridge and the other one with a elastic net penalty. Recall that ridge regression uses l_2 penalty and yields highly polygenic predictors (all non zero weights) whereas elastic net yields sparse models, setting the weights of most variants to 0. Given the high polygenicity of IDPs, we expected ridge regression to perform better than the sparsity-inducing elastic net penalty (Methods).

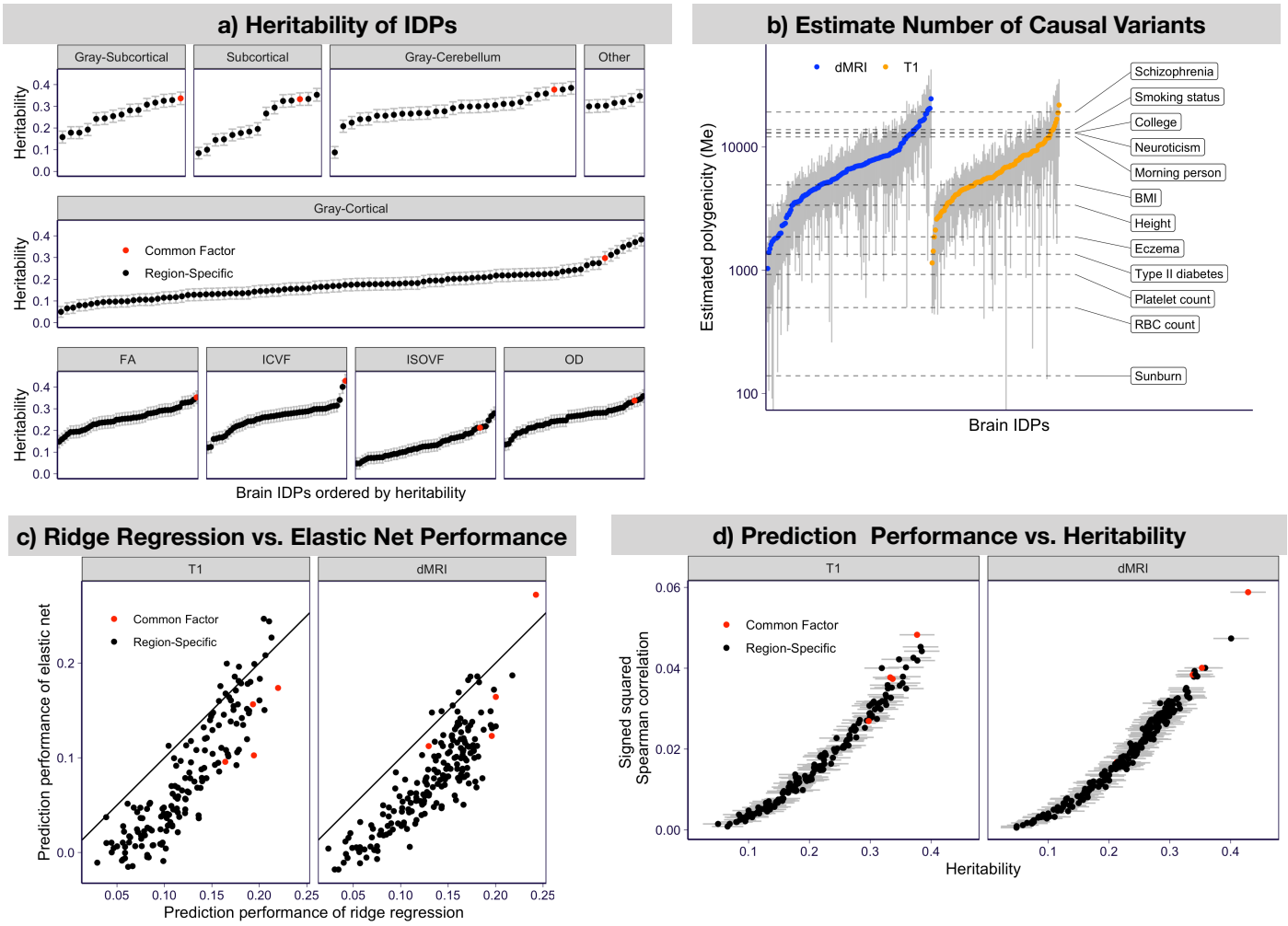


Figure 3: Genetic architecture of IDPs and prediction. **a) Heritability of IDPs.** Heritability and 95% confidence intervals are shown by IDP subtype: volumes of gray matter subcortical structures, total volume of subcortical structures, volume of cortical regions and other volumes among structural IDPs. The third row show heritability of diffusion IDPs. Red show common factors (principal components of each subtype) and black are the component IDPs in the subtype. FA: fractional anisotropy; ICVF: intracellular volume fraction; ISOVF: isotropic volume fraction; OD: orientation dispersion index. **b) Polygenicity of IDPs.** The effective number of independently association SNPs (M_e) estimated using the stratified LD fourth moments regression (O'Connor et al., 2019) for 522 brain IDPs are shown on the y-axis. Gray bars indicate 95% confidence intervals. Horizontal dashed lines indicate the estimated M_e of 12 complex traits for reference. **c) Prediction performance of IDPs.** This panel compares the performance of ridge regression vs. elastic net approaches predictors measured by the Spearman correlation between the predicted and observed IDP values in a five-fold cross-validated scheme. For each brain IDP, the performance of ridge predictor is shown on the x-axis and the performance of elastic net predictor is shown on the y-axis. The black solid line is the identity line ($y = x$). The IDP PCs are in red and the remaining brain IDPs are in black. **d) Prediction performance vs. heritability.** The signed squared Spearman correlations between observed and predicted IDPs of the ridge regression are shown on y-axis and the estimated heritability are shown on x-axis. The signed squared correlation is defined as $\text{sign}(x) \cdot x^2$ for correlation x to preserve the sign of the correlation while taking the square. The error bar indicates the 95% confidence interval of the estimated heritability. The IDP PCs are in red and the rest of the brain IDPs are in black.

To evaluate the model performance we calculated the Spearman correlation between the observed and predicted IDP values with a five-fold cross-validation scheme (Methods), shown in Fig. 3c. For ridge regression the prediction performance ranged from 0.024 to 0.24 with a median of 0.13. For elastic net the range was between -0.018 and 0.27 with a median of 0.075. For most IDPs, ridge regression yielded higher cross-validated performance than elastic net (See Fig. 3c.)

We also found that the improved performance of ridge regression over elastic net predictions was correlated with the polygenicity (M_e) of the trait (fig. S6), corroborating our intuition that ridge regression performs better for polygenic traits whereas elastic net performs better for more sparse traits.

All the ridge predictors and 95% of the elastic net predictors showed positive cross-validated Spearman correlation (fig. S7 and table S2) demonstrating the feasibility of genetically predicting IDPs. With these predictors in hand we moved to the next stage of the development of the framework where we predicted IDPs using genotype alone and correlated these predicted values with complex traits.

As expected, the prediction performance increased with the heritability of the IDP, i.e. more heritable traits were predicted better as shown in Fig. 3d. However, we also noted that the median prediction R^2 was lower than 8% of the heritability (Fig. 3c), an upper bound of how the performance of linear predictors. This low proportion of heritability captured by the genetic predictors highlights the need to increase the sample size of reference image data to reach the upper bound of the performance.

We filtered out unreliable predictors by keeping only the ones that showed prediction performance correlation greater than 0.1. Among structural IDP residuals, 105 ridge predictors and 54 elastic net predictors passed the threshold from a total of 159 trained. Among 192 diffusion IDP residuals, 148 ridge predictors and 62 elastic net predictors passed the threshold. All subtype-level PCs except the elastic net-based PC predictor of cortical region volumes were well predicted and kept for the subsequent analysis.

Summary BrainXcan finds disease-associated brain features using GWAS summary statistics

To expand the applicability of our tool, we extended the BrainXcan association module so that it could infer the association statistics using the GWAS summary results of the traits without the need to use individual level genotype and phenotype data. The ideas are similar to the S-PrediXcan method used for correlating genetically predicted transcriptome with complex traits. However, unlike gene expression prediction which only used variants in the vicinity of the gene, IDPs needed to handle a dense number of genetic predictors across the genome. To make this computation feasible, we developed a scalable method that could handle this added complexity as described in the Method section.

To enable the application of the method to situations where only summary statistics are available, we calculated the covariance between the HapMap3 SNPs to be used downstream and saved in a sparse format, setting correlations between SNPs that were separated by more than 200 SNPs to be 0. We also performed genetic associations of all the IDPs (GWAS of IDPs) and saved the results for the Mendelian randomization analysis downstream.

BrainXcan association: correlating genetically predicted IDPs with phenotypes

We selected 9 phenotypes from the UK Biobank and performed BrainXcan association using data from 327,743 individuals of British ancestry. As described in the overview section above, we calculated the genetically predicted IDPs for all the individuals and correlated them with the phenotypes using linear regression. The phenotypes included alcohol consumption, smoking, coffee consumption, depression, parental depression, parental Alzheimer's disease, handedness, BMI, and height. See detailed list on table S3. To avoid overfitting issues, we excluded individuals used for the training of the prediction models.

We also performed summary BrainXcan association analysis on 35 GWAS for which we did not have access to the individual level data. These phenotypes included behavioral, psychiatric and neurologic phenotypes, height, and body mass index (see table S4).

We confirmed the reliability of the summary version of BrainXcan (S-BrainXcan) by comparing the z-scores of the associations to the ones obtained from individual-level BrainXcan for standing height and body mass index in the UK Biobank. The highly concordant z-scores are shown in fig. S11. We also observed concordant BrainXcan z-scores (fig. S8) between ridge and elastic net predictors indicating robustness to the choice of prediction approach. Ridge predictors yielded more significant results, consistent with their increased prediction performance compared to elastic net.

Combining both summary level and UK Biobank traits, 98% of IDPs (257 out of the 261 IDPs in the main analysis) were significantly associated with at least one trait. As expected, better powered GWAS traits with larger number of significant associations also yielded more significant IDPs to trait associations.

Better predicted IDPs were more significantly associated with phenotypes, likely due to reduced attenuation bias (fig. S9). Brain-wide features represented by PC's were more significantly associated with complex traits than region-specific features which could be explained by the higher predictability of the PC's but could also point to brain-wide effects being more prevalent among the selected GWAS traits (fig. S10).

Association results replicate in independent datasets

For a number of traits, we had two independent GWAS of the same trait, which allowed us test the replication of our results. Reassuringly, as shown in Fig. 4a, we found highly concordant association z-scores between

the independent studies.

Genetic correlations yield similar but less significant associations

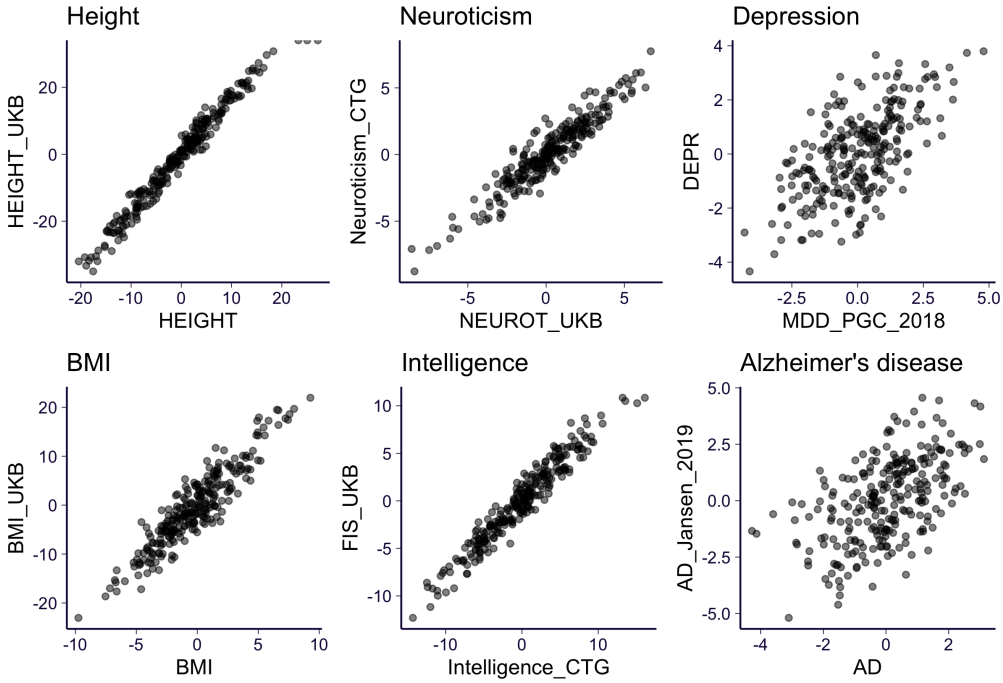
Genetic correlation between brain features and complex traits can provide, in principle, similar information to BrainXcan association results. To compare the ability to identify associations, we computed the correlations for all pairs of IDP/complex traits and compared the results to BrainXcan associations. (Methods and Figure 4b). We found that the z-scores of the genetic correlation was highly correlated to the z-scores of BrainXcan associations (with correlations ranging from 0.51 to 0.97, with a median of 0.81 (fig. S13). However, BrainXcan yielded 8-fold larger number of significant IDP/trait associations compared to genetic correlation approach suggesting that optimal predictors of IDPs yield more power to identify putative causal links.

BrainXcan quantifies evidence for the direction of the causal flow

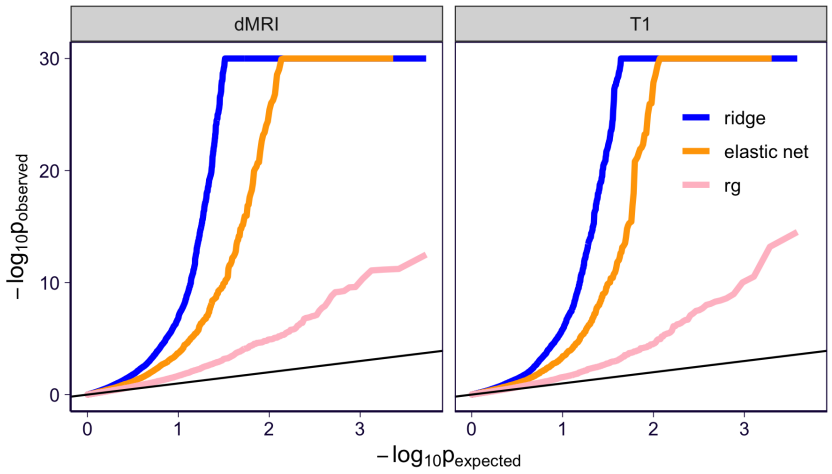
Mendelian randomization evaluates the causal relationship between trait 1 and 2 by testing whether increased "exposure" to the first trait (represented as the trait 1 GWAS effect size) is associated with increased or decreased level of the second trait (indicated by the trait 2 effect size). In Mendelian randomization settings, individuals are thought to be randomized at meiosis to either inherit the risk increasing allele or not. Because of this parallel to randomized trials, the level of causal evidence derived from Mendelian randomization is considered to be higher than observational studies albeit lower than actual randomized trials (the gold standard for causal determination used in clinical trials).

It is possible to infer the direction of the causal flow by selecting variants that have strong effects on the first trait and testing for a significant association with the effect sizes of the second trait and vice versa, i.e. selecting the variants with strong effects on the second trait and testing whether they are associated with the effect sizes on the first trait. As described below, scatter plots of effect sizes showing the two selection strategies (by significance of trait 1 or 2), are added to the automated reports.

To determine the direction of the putative causal flow, i.e. whether changes in IDPs are affecting changes in the trait or vice versa, we applied a number of Mendelian randomization approaches with multiple instruments implemented in MR-BASE (Hemani et al., 2018) including inverse variance weighted regression (Burgess et al., 2013), weighted median method (Bowden et al., 2016), and Egger regression (Bowden et al., 2015). For this purpose, we selected strong instruments, i.e. SNPs with very small p-values. For the complex phenotype, we selected SNPs with GWAS p-values $< 5 \times 10^{-8}$ and for brain IDPs, we selected SNPs with GWAS p-values $< 10^{-5}$. The more loose p-value for IDPs were necessary to have sufficient number of instruments, which could be tighten as the number of individuals in the reference image data increases. To



(a)



(b)

Figure 4: S-BrainXcan association statistics for 35 GWAS. (a) Six phenotypes which has multiple GWASs being analyzed by S-BrainXcan are shown. For each brain IDPs, the S-BrainXcan z-scores (from ridge models) from the two GWASs are shown on x-axis and y-axis respectively (see the GWAS label in table S4). **(b)** The S-BrainXcan p-values are shown as the QQ-plot (against the expected p-values under the null). For visualization purpose, the observed p-values which is smaller than 1×10^{-30} are set to 1×10^{-30}. Label 'rg' represents the genetic correlation result. The black lines are the identity line ($y = x$).

streamline the interpretation of the multiple Mendelian randomization results, we combined the p-values of each Mendelian randomization output using an extension of the ACAT method (Liu et al., 2019) that take into account the concordance of the sign of the results.

Caveats on interpreting Mendelian randomization results

There are two caveats that need to be considered when interpreting Mendelian randomization results. One is that we first select the IDP-trait pair based on their association. Therefore, the p-values of the Mendelian randomization will not be well-calibrated. Therefore, we propose using the p-values to discern between possible direction of the causal flow, not to quantify significance.

The second caveat relate to the power difference between reference image and GWAS studies. Currently, reference image data have much smaller sample sizes (n$\sim 30K$) compared to GWAS studies of complex traits (n$\sim 100K - 1M$). We consider three main mediating scenarios depicted in Fig. 2b). In scenario i) the brain feature mediates the genetic association with the phenotype, i.e. genetic risk factors alter the brain feature which in turn alters the risk for the phenotype. In scenario ii) genetic factors affect the phenotype which in turn alter the brain feature. In scenario iii) genetic factors affect an underlying latent factor which alter both the phenotype and the brain feature.

A significant i) and not significant ii) can be interpreted as evidence that the brain feature alteration is affecting the phenotype given the higher power of GWAS studies in general. However, a significant scenario ii) and non significant scenario i) could simply mean that the instruments (strongly associated SNPs and their effect sizes) for the brain feature are not reliable enough to yield significance. In this case, scenario ii) should be considered supported by the data but scenarios i) and iii) should not be ruled out.

BrainXcan use is simplified with an automated pipeline

To facilitate the analysis to first-time users, we created an automated pipeline using snakemake (Mölder et al., 2021). The association and MR modules can be performed with default parameters or modified according to the study's specific needs. The pipeline tests all the IDPs for associations and Mendelian randomization test is performed for the top 10 (modifiable parameter of the pipeline). The pipeline will

- run BrainXcan association ($Y \sim$ resIDP and $Y \sim$ PC),
- run bidirectional Mendelian randomization for the top 10 significant IDPs, and
- generate automated report including figures, tables with top associations, Mendelian randomization figures, etc.

Application of BrainXcan to Schizophrenia

To demonstrate the features of BrainXcan, we applied the full pipeline to the schizophrenia GWAS (Ripke et al., 2020). See the automated report in the Supplementary Note. In this analysis, we tested 327 structural and diffusion MRI-derived phenotypes with cross validated Spearman correlation greater than 10%. For the main analysis we focused on 261 IDPs, which include 48 cortical gray matter volumes, 10 subcortical volumes, 13 subcortical gray matter volumes, fractional anisotropy (water diffusivity along nerve tracts) in 46 regions, ICVF (intracellular volume fraction) in 44 regions, OD (orientation dispersion) in 45 regions, and ISOVF (isotropic volume fraction) in 13 regions. (Note that ISOVF was a less heritable index leading to fewer successful predictors). Brain-wide measures represented by principal components of each subtype were also included (gray volumes of cortical, cerebellum, and subcortical regions, subcortical total volumes, FA, ICVF, OD, and ISOVF).

Among the 261 IDPs, 46 were significantly associated with risk of schizophrenia after Bonferroni correction (raw p-value $< 0.05/261$). Figs. 7 and 8 provide a snapshot of the region-specific associations schizophrenia risk. We added an interactive annotation of different regions of the brain is added to the output of the automated pipeline. See example in `https://brainxcan.hakyimlab.org/post/2021/05/06/brainxcan-automated-reports/#interactive-annotation-of-regions`

Among diffusion MRI associations, the principal component of ICVF, a proxy for brain-wide axonal density, was the most significant association for schizophrenia risk, with lower axonal density associated with higher risk of schizophrenia (Fig. 5). The principal component of fractional anisotropy, a brain-wide measure of water diffusion efficiency along nerve tracts, was also negatively associated with schizophrenia while the orientation dispersion index (dispersion of neurite orientation along tracts) did not show significant association. These results corroborate the hypothesis that schizophrenia is a disorder of disconnectivity. They are also consistent with reduced FA in schizophrenia cases compared to controls as reported by Kelly et al. (2018). However, since our technique uses healthy individuals MRI-based genetic prediction of brain features, they are less likely to capture a consequence of the disease.

The total volume of the hippocampus (right side, relative to brain size) was positively associated with schizophrenia risk whereas the total volume of the thalamus (both sides, relative to brain size) were negatively associated. Gray matter volumes of the frontal orbital cortex, the anterior cingulate cortex, and posterior temporal fusiform cortex were positively associated whereas planum polare, amygdala's gray matter volumes were negatively associated (Fig. 6). Observed hippocampus, thalamus, amygdala, anterior cingular cortex volumes have all been reported to be associated with schizophrenia status (van Erp et al., 2016; Shepherd et al., 2012).

Among the top 10 IDPs associated with schizophrenia, we found that 2 IDP to schizophrenia risk causal

flow were nominally significant ($p < 0.05$) and 5 schizophrenia to IDP causal flow were nominally significant. Notice that the IDPs used for the training of the prediction models did not have schizophrenia. Therefore, our design cannot lead to scenario ii in Fig. 2b which needs the individuals in the IDP prediction training set to have been exposed to the schizophrenia. Given the caveats discussed in the Results section and the fact that ii) cannot be detected in our design, we conclude that our results provide support for scenario iii).

Discussion

We propose a robust and scalable framework we call BrainXcan, which leverages genetically predicted brain features trained in reference MRI datasets, genome-wide association studies of complex diseases, and computational and methodological advances to dissect the biology of behavioral, neurological, and psychiatric traits. Our approach addresses the sample size limitations of MRI studies by taking advantage of increasing cohorts of GWAS studies and large MRI data in predominantly healthy subjects. The use of genetic variation helps us circumvent the reverse causality problem.

Our association module identifies brain features likely to have an effect on behavioral and psychiatric traits but also features that can be modified by the disease. Our Mendelian randomization module quantifies the evidence for each of the direction of the putative causal flow (brain feature to disease or vice-versa). Naturally, both direction of the effects are informative. Understanding how human disease cause changes in brain features captured by MRI can help design better diagnostic tools. Brain features that modulate the risk to disease provides insights into the pathogenesis and can help identify preventive or therapeutic strategies.

To encourage broad adoption of our tools by users less familiar with genetic toosl, we provide a user friendly pipeline and an automated report. All the tools necessary to perform prediction, association, and causal flow assessment is provided (`https://github.com/hakyimlab/brainxcan`).

In the process of developing the tools, we learned that brain features are highly polygenic, in many cases even higher than human height (the canonical example of a polygenic trait), more similar to psychiatric and behavioral traits. The similarity in genetic architecture suggests that brain features can be useful endo-phenotypes to improve the classification of complex psychiatric diseases.

We present an application to schizophrenia to showcase the potential of our method. The significant association between low levels of brain-wide ICVF, proxy for axonal density, with high risk of schizophrenia corroborates the long standing disconnectivity hypothesis of schizophrenia. We also find the volumes of many regions of relevance associated with schizophrenia risk, including the amygdala, hippocampus, and

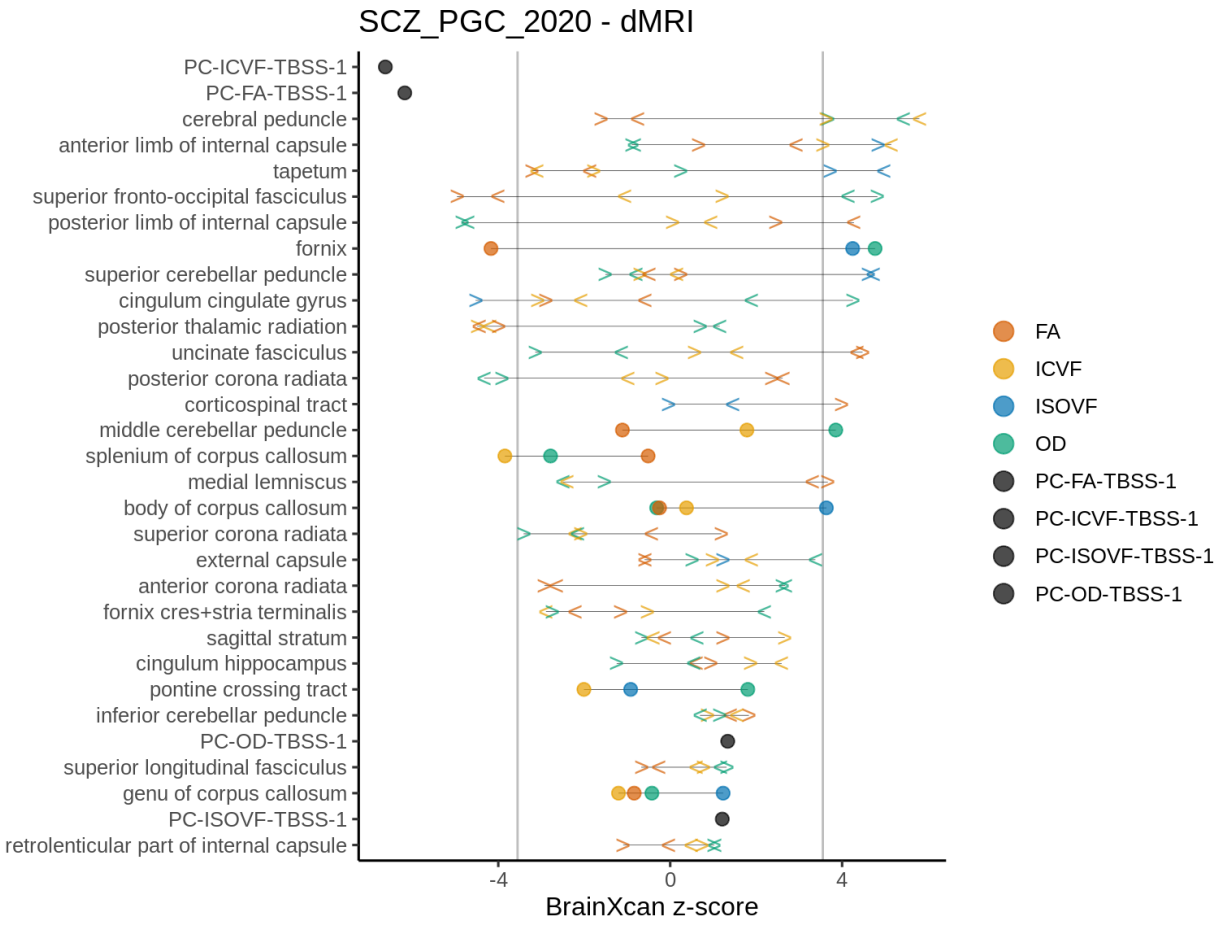


Figure 5: Schizophrenia risk association with diffusion MRI. Z-scores of region-specific associations of IDP with schizophrenia risk using GWAS effect sizes reported in Ripke et al. (2020). The features starting with "PC" correspond to the principal components of IDP across regions and are proxies for the brain-wide feature for the subtype. FA: fractional anisotropy, ICVF: intracellular volume fraction, ISOVF: isotropic volume fraction, OD: orientation dispersion index. < indicates left, > indicates right, circles are used when sides are not defined

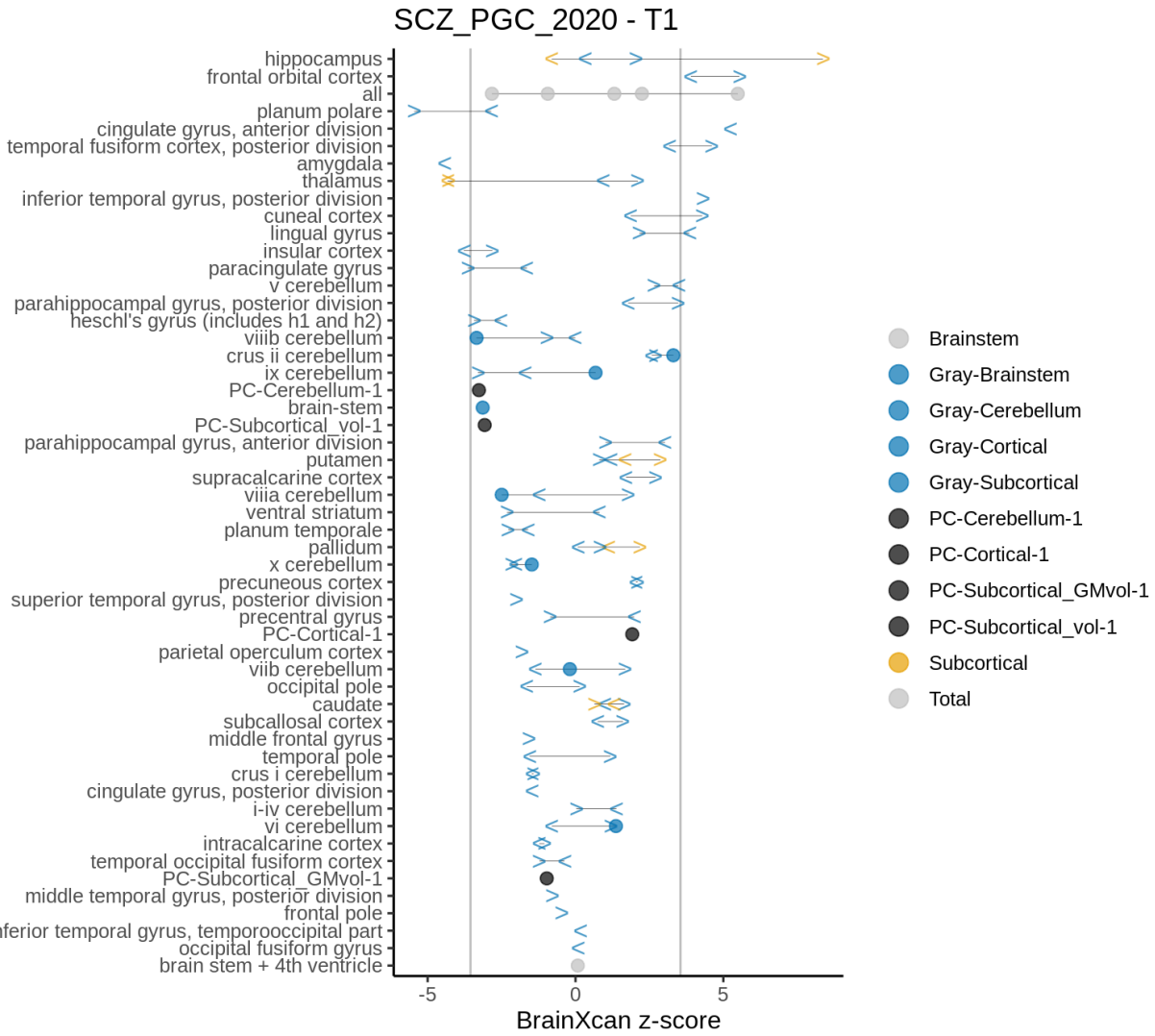


Figure 6: Schizophrenia risk association with structural features. Z-scores of region-specific association of IDP with schizophrenia risk using GWAS effect sizes reported in Ripke et al. (2020). The features starting with "PC" correspond to the principal components of IDP across regions and are proxies for the brain-wide feature for the subtype. In blue are shown the gray matter volumes of the cortex, subcortex, brainstem, and cerebellum. In black are shown the principal components of each category. In yellow are shown the associations with subcortical total volumes quantified with FIRST. < indicates left, > indicates right, circles are used when sides are not defined.

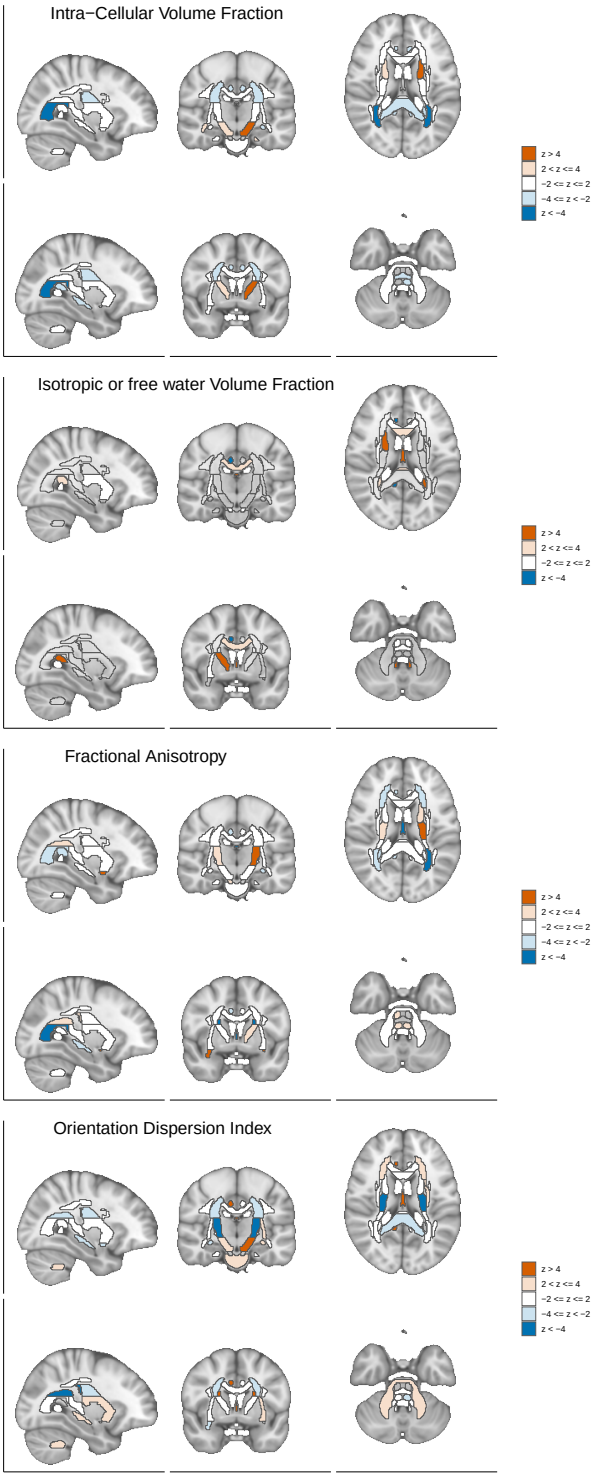


Figure 7: Brain visualization of diffusion features associations with schizophrenia risk. Z-scores of the associations between the brain region and schizophrenia risk are shown with different slices of the brain.

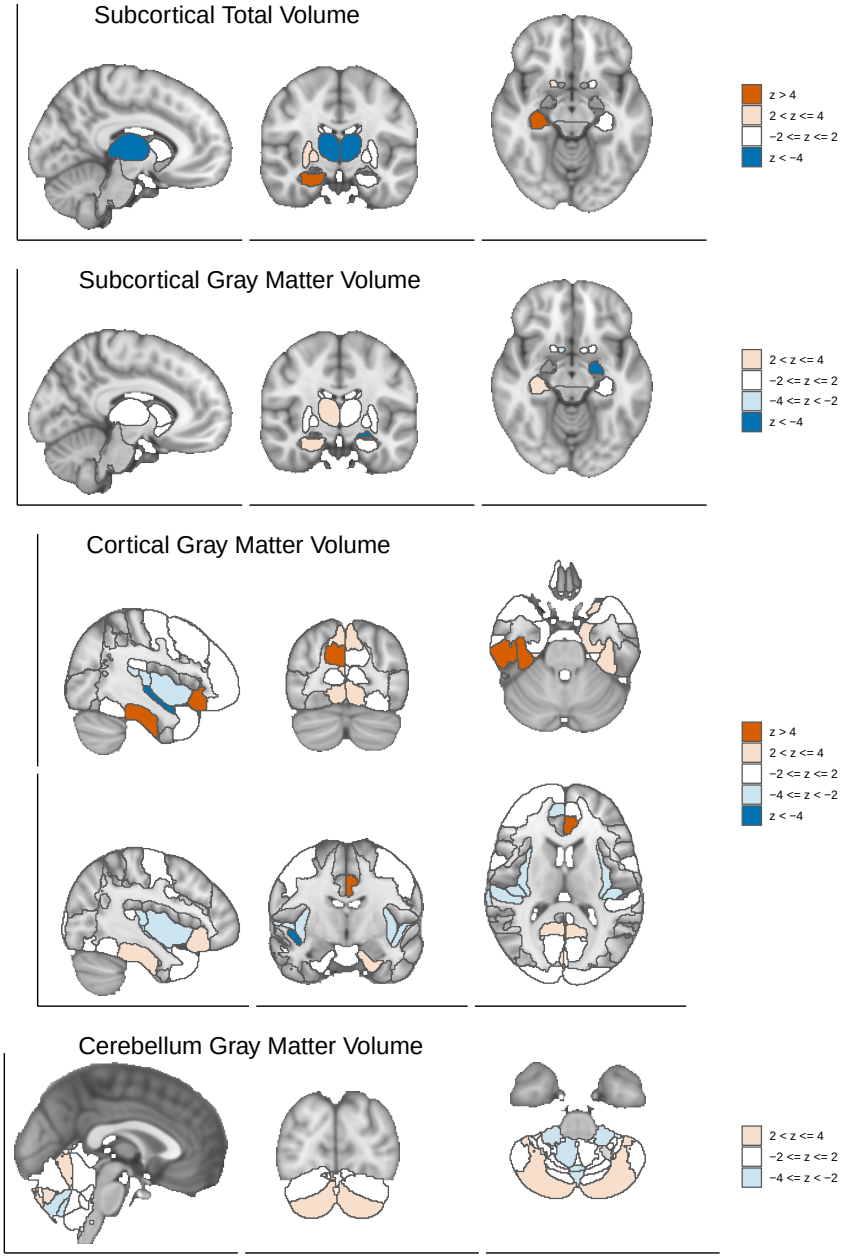


Figure 8: Brain visualization of structural features associations with schizophrenia risk. Z-scores of the associations between the brain region and schizophrenia risk are shown with different slices of the brain.

the anterior cingular cortex. Our results add robust orthogonal lines of evidence to existing literature since our associations are based on the genetic components of the traits and are less likely to be confounded than studies with observed traits.

We provide a user-friendly software package to attract investigators less familiar with the genetic field. Our software is implemented in R and python with a streamlined pipeline coded in snakemake. An automated report in html format with pre-defined figures summarizing the results facilitates interpretation.

We note that there are several limitations in the current work. First, the prediction performance of the current genetic predictors are largely limited by the size of the training cohort (Fig. 3d). As the UK Biobank is gradually collecting more brain imaging data (Littlejohns et al., 2020), we expect the training cohort size will increase to at least 100,000 individuals. We will be updating the the brain IDP predictors as the data becomes more available. Second, the S-BrainXcan calculation relies on the genotype covariance which is approximated as a banded matrix (Methods) here. This approximation may, particularly, affect the stability of the joint analysis since the jointly analysis relies heavily on the predicted IDP covariance which is derived from the genotype covariance. This was one of the reasons we decided not to pursue the joint model since they can lead to false positives. Third, the BrainXcan analysis cannot establish the causal relation between the brain IDPs and the complex phenotype. Although we run the Mendelian randomization in both the forward and the reverse directions for the IDP/phenotype candidate, the Mendelian randomization results should be interpreted with caution due to the following reasons: i) different Mendelian randomization tests may not give consistently significant results; ii) the Mendelian randomizations of the forward and the reverse directions typically have different power; iii) if the same GWAS is used for both BrainXcan association and Mendelian randomization, the Mendelian randomization p-value is not well-calibrated; iv) causality is valid only when the Mendelian randomization assumptions are hold. Given these limitations, we consider the Mendelian randomization results to be quantification of the evidence for the direction of the putative causal flow. Fourth, only linear prediction models are used in our implementation. More sophisticated models could be used for prediction but that will limit the application to cases where the full individual level data is available. Given the current availability of summary results and the need to use very large sample sizes to reliably estimate genetic effects makes the use of non-linear models less than optimal. Despite these limitations, we anticipate that BrainXcan, a user-friendly analysis tool, will be broadly adopted and help in identifying brain features important in the pathogenesis as well as diagnosis of complex phenotypes.

Methods

Preprocessing of UK Biobank IDP phenotypes

We queried from UK Biobank database using ukbREST (Pividori and Im, 2019) to retrieve the list of 459 IDP phenotypes as shown in table S1 (Smith et al., 2020). Among these IDPs, 400 IDPs are diffusion MRI measurements including 192 TBSS-style measurements (TBSS) and 108 probabilistic-tractography-based measurements (ProbTrack). The remaining 159 IDPs are T1-weighted structural measurements, including 139 FAST-based grey matter segmentation based measurements and 14 FIRST-based subcortical structures measurements and 6 T1 structural brain MRIs measuring the total volumes of the peripheral cortical grey matter, ventricular cerebrospinal fluid, brain grey matter, brain white matter, brain grey + white matter, and brain stem. In total, we collected 24,409 European-descent individuals in UK Biobank with non-missing IDPs and non-missing values for other covariates such as genetic PCs, sex, and age at recruitment.

We scaled the structural IDP's using the volumetric scaling factor from T1 head image (UK Biobank Data-Field 25000) so that the measurement of the brain region volume was relative to the total brain volume. We regressed out the following scanner position covariates of IDPs: UK Biobank data rields 25756, 25757, 25758, and 25759. We also regressed out the first 10 genetic PCs, age, sex, squared age, age $\times$ sex, and squared age $\times$ sex.

To adjust for the correlation between IDPs and to extract the common factor of IDPs, we performed principal component analysis on the IDP matrices (individual-by-IDP matrices). Here, the PCA was done for each IDP subtype. For T1 IDPs, the subtypes were gray matter volume of the cortical regions, gray matter volume of the subcortical regions, total volume of subcortical regions, and gray matter of the cerebellum regions. For dMRI IDPs, the subtypes were defined by the four measure types (FA, ICVF, ISOVF, and OD) of TBSS and ProbTrack IDPs respectively. We obtained the first PC as the measure of the common factor for each IDP subtype. We regressed out the first PC from the IDPs and obtained the IDP residuals. Finally, we inverse-normalized the PC-adjusted IDP residuals and also the IDP PCs for the subsequent analysis. We refer the IDP residuals and IDP PCs as brain IDP's.

Selecting variants from UK Biobank imputed genotypes

We extracted the list of common variants (minor allele frequency > 0.01) among CEU individuals in HapMap 3 data (International HapMap 3 Consortium and others, 2010). And then, we excluded ambiguous variants which have reversely complementary bases as the reference and the alternative alleles (AT and CG pairs). In total 1,078,323 variants passed the criteria and among these, 1,071,650 variants appeared in the UK Biobank imputed genotype data (by SNP rs ID). In the subsequent analysis, we limited the computations to this set

of variants.

Estimating the heritability

We estimated the heritability of IDPs assuming random effects for SNPs (Yang et al., 2010). We used the EMMA algorithm proposed in Kang et al. (2008) to avoid the repeated calculations when dealing with multiple phenotypes on the same cohort. In the analysis, the genetic relatedness matrix (GRM) was built using the pre-selected HapMap 3 SNPs with minor allele frequency > 0.05. Since we accounted for the covariates in the preprocessing steps, we included no covariates other than the intercept in the heritability calculation.

Estimation of polygenicity

We estimated the effective number of independently associated SNPs using the stratified LD fourth moments regression method (O'Connor et al., 2019). We downloaded the pre-computed LD scores and LD fourth moments from `https://www.dropbox.com/sh/iiyftw01gdpt6un/AACU7AmWK45RxTmDJvRkdKhIa?dl=0` and used the scores stored in `baselineLD.1kg.l2l4.mat`. These scores were based on baselineLD annotations (Gazal et al., 2017). The stratified LD fourth moments regression was performed in MATLAB by calling `SLD4M` function shared in `https://github.com/lukejoconnor/SLD4M` (O'Connor et al., 2019). To obtain the effective number of independently associated common SNPs, we aggregated the estimated M_e values across 10 MAF bins which correspond to the common variants (MAF > 0.05). The aggregation was done by setting `report_annot_mat` variable accordingly inside the `SLD4M` function.

Building polygenic predictors for IDPs

We built polygenic predictors using both ridge regression models and elastic net models for the brain IDPs. The models were fitted using the pre-selected 1,078,323 genome-wide variants described above. Since covariates had already been adjusted for, we included only the intercept as covariate in the model training.

Ridge regression models

In the ridge regression, we fit the following optimization problem:

$$\arg\min_{w} \|B - Xw\|_2^2 + \lambda\|w\|_2^2, \tag{3}$$

where B is the mean-centered brain IDP (fitting one IDP at a time), X is the standardized genotype matrix of the variants, and w is the prediction model weights. To reduce the computation complexity and take advantage of the fact that the number of variants is much larger than the number of samples, we used the formula similar to the one in OmicKriging (Wheeler et al., 2014). Specifically, the ridge regression

estimator $\widehat{w}^{\text{ridge}} = (X'X + \lambda I)^{-1}X'B$ (where λ is the hyperparameter) involves solving a linear system with $P \times P$ matrix where P is the number of variants. Alternatively, it can be re-arranged as $\widehat{w}^{\text{ridge}} = X'(XX' + \lambda I)^{-1}B$ instead and, correspondingly, the predicted value of B is $\widehat{B} = (XX')(XX' + \lambda)^{-1}B$. Let Σ represent the GRM matrix and we have $\Sigma = XX'/P$. The expression for the $\widehat{B}$ can be re-parameterized as $\widehat{B} = \theta\Sigma^{-1}(\theta\Sigma + (1-\theta)I)^{-1}B$ with $\theta = \frac{P}{P+\lambda} \in (0, 1]$. We performed 5-fold cross-validation on a grid of θ values 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 to choose the hyperparameter θ.

Elastic net models

The elastic net models were fitted using the R package `snpnet` which implements the BSAIL algorithm proposed in Qian et al. (2020). Specifically, we fit the following optimization problem:

$$\arg\min_{w} \frac{1}{2N}\|B - Xw\|_2^2 + \lambda(\frac{1-\alpha}{2}\|w\|_2^2 + \alpha\|w\|_1) \tag{4}$$

We fixed the hyperparameter $\alpha = 0.1$ and determined the value of λ by the R^2 on the validation set. First, we randomly split the data into two parts: the training set (containing 80% of the samples) and the validation set (containing 20% of the samples). Then, using the training set, we trained a series of elastic net models under a grid of λ values. All these models were evaluated on the validation set by calculating the R^2 between the observed and predicted B. We selected the λ with the largest validation R^2 and trained the final model using the full data (combining the training and validation sets).

When using `snpnet`, we set the maximum number of iterations (`niter`) to 100 and the number of SNPs to consider in each batch (`num.snps.batch`) to 200. The phenotypes were fitted one at a time.

Calculating the prediction performance

For both ridge and elastic net models, we evaluated the prediction performance by 5-fold cross validation. At each split, we used the 4 folds as if the full data and trained the models using the procedures described above. And then, we made the prediction on the held-out one fold of the data. This procedure was repeated for all five splits and the predicted values were aggregated across all folds. The prediction accuracy was evaluated in terms of R^2, Pearson correlation, and Spearman correlation.

BrainXcan with individual-level data

We performed individual-level BrainXcan as described above on a set of 327,743 unrelated, European-descent UK Biobank participants who were not included in the brain IDP model training. We included sex, age, squared age, age $\times$ sex, squared age $\times$ sex, and the first 10 genetic PCs as covariates.

BrainXcan with summary statistics

When the individual-level information is not available, we calculated the BrainXcan statistic approximately using the GWAS summary statistic and the genotype covariance from a reference panel. The formulas are similar to the ones proposed in Barbeira et al. (2018). Let $\widehat{b}_j$ and $\mathrm{se}(\widehat{b}_j)$ be the estimated effect size and the corresponding standard error for variant j from the GWAS. And $z_{\mathrm{GWAS},j}$ is the GWAS z-score of SNP j. Let $\widehat{R}$ represent the genotype sample covariance matrix where $\widehat{R}_{jj'} = \widehat{\mathrm{Cov}}(X_j, X_{j'})$, namely the sample covariance between variant j and j'. We can calculate the marginal test statistics using the following results (see derivations in Supplementary Notes 3):

$$\hat{\beta}_k = \frac{\sum_{j=1}^{P} \widehat{w}_{kj} \widehat{R}_{jj} \widehat{b}_j}{\widehat{\sigma}_k^2} \tag{5}$$

$$z_{\mathrm{BrainXcan},k} \approx \frac{\sum_{j=1}^{P} \widehat{w}_{kj} \sqrt{\widehat{R}_{jj}} z_{\mathrm{GWAS},j}}{\widehat{\sigma}_k} \tag{6}$$

$$\widehat{\sigma}_k^2 = \sum_{j=1}^{P} \sum_{j'=1}^{P} \widehat{w}_{kj} \widehat{w}_{kj'} \widehat{R}_{jj'}, \tag{7}$$

where $z_{\mathrm{BrainXcan},k}$ represents the BrainXcan marginal test z-score for the kth brain IDP. We refer this summary statistic-based BrainXcan as S-BrainXcan.

In principle, we need to consider the genotype covariance for all the genome-wide variant pairs. In this paper, to reduce the computation burden, we first assumed that the between-chromosome covariance is zero. Moreover, we considered the per-chromosome genotype covariance matrix as a banded matrix with bandwidth equal to 200. In other words, any variant pairs with more than 200 variants in-between are considered having zero covariance. We used the set of 24,409 UK Biobank individuals used for the brain IDP model training as the reference panel for genotype covariance calculation.

Performing GWAS for brain IDPs

We performed genome-wide association studies for all the brain IDPs using Python package `tensorqtl` (Taylor-Weiner et al., 2019). Since we adjusted covariates in the preprocessing of the brain IDPs (see previous sections), we did not include any covariates other than the intercept in the GWAS runs.

Mendelian randomization analysis of IDP/phenotype pairs

To investigate the direction of the effect, we performed Mendelian randomization (MR) analysis for the significant IDP/phenotype pairs identified in the BrainXcan association stage. The MR analysis was performed for both directions: i) brain IDP $\rightarrow$ phenotype, treating the brain IDP as the mediating trait and the phenotype of interest as the outcome trait; ii) phenotype $\rightarrow$ brain IDP, the phenotype of interest as the

mediating trait and the brain IDP as the outcome trait. As the inputs of the analysis, we used the brain IDP GWAS results as described in the above section. For the phenotype of interest, we used the GWAS results which were also used for the S-BrainXcan analysis.

For the mediating trait, the instrument variants were selected using LD clumping function (`ld_clump`) in the R package `ieugwasr` (Elsworth et al., 2020). We used the EUR super-population in 1000 Genomes data (1000 Genomes Project Consortium, 2015) as the LD reference panel and the data was downloaded from `http://fileserve.mrcieu.ac.uk/ld/1kg.v3.tgz`. The LD clumping parameters were `clump_kb = 10000` and `clump_r2 = 0.001`. The p-value parameter (`clump_p`) in the LD clumping was 10^{-5} for IDP GWAS and 5×10^{-8} for phenotype GWAS, which gave approximately independent and significant variant instruments.

The MR analysis was performed using the R package `TwoSampleMR` (Hemani et al., 2018). We reported the MR results using three MR methods: i) inverse variance weighted MR (Burgess et al., 2013); ii) median-based estimator: weighted median (Bowden et al., 2016); iii) MR Egger analysis (Bowden et al., 2015), which corresponds to `mr_ivw` , `mr_weighted_median`, and `mr_egger_regression` in `TwoSampleMR`. We further reported a meta-analyzed p-value summarizing the results of the three MR tests being performed. The meta-analysis is based on an extension of ACAT method (Liu et al., 2019) that takes into account the direction of the effects. See Supplementary Notes 4 and fig. S14 for additional detail.

Calculating the genetic correlation for IDP/phenotype pairs

The genetic correlation between a brain IDP and the phenotype of interest was calculated using the cross-trait LD Score regression (Bulik-Sullivan et al., 2015). The actual calculation was performed using the Python package `ldsc` (`https://github.com/bulik/ldsc`). The pre-computed LD-scores were downloaded from `https://storage.googleapis.com/broad-alkesgroup-public/LDSCORE/eur_w_ld_chr.tar.bz2` which are based on the 1000 Genomes European data. We used the brain IDP GWAS results as described in the above section and the GWAS of the phenotype was the same as the one used for the S-BrainXcan analysis.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 19526. YL, OM, and HKI received partial funding from P30DK020595, U01HG009086.

Code and data availability

`https://github.com/hakyimlab/brainxcan`

Contributions

Y.L. designed and built the computational pipelines, performed the data analysis, created the software and visualization for BrainXcan, and wrote the original draft of the manuscript. O.M. performed initial data cleaning and exploration and contributed to the pipeline development. T.B provided the computational resources and assisted the computation. A.B discussed and interpreted the analysis results. H.K.I. supervised the whole project, designed the computational pipelines and data analysis, and extensively edited the manuscript. All authors read and approved the final manuscript.

References

1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526(7571):68–74, 2015.

A. N. Barbeira, S. P. Dickinson, R. Bonazzola, J. Zheng, H. E. Wheeler, J. M. Torres, E. S. Torstenson, K. P. Shah, T. Garcia, T. L. Edwards, others, and H. K. Im. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from gwas summary statistics. *Nature communications*, 9(1):1–20, 2018.

J. Bowden, G. Davey Smith, and S. Burgess. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *International journal of epidemiology*, 44(2): 512–525, 2015.

J. Bowden, G. Davey Smith, P. C. Haycock, and S. Burgess. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic epidemiology*, 40(4): 304–314, 2016.

B. Bulik-Sullivan, H. K. Finucane, V. Anttila, A. Gusev, F. R. Day, P.-R. Loh, L. Duncan, J. R. Perry, N. Patterson, E. B. Robinson, et al. An atlas of genetic correlations across human diseases and traits. *Nature genetics*, 47(11):1236, 2015.

S. Burgess, A. Butterworth, and S. G. Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology*, 37(7):658–665, 2013.

A. Dahl, V. Guillemot, J. Mefford, H. Aschard, and N. Zaitlen. Adjusting for principal components of molecular phenotypes induces replicating false positives. *Genetics*, 211(4):1179–1189, 2019.

L. T. Elliott, K. Sharp, F. Alfaro-Almagro, S. Shi, K. L. Miller, G. Douaud, J. Marchini, and S. M. Smith. Genome-wide association studies of brain imaging phenotypes in uk biobank. *Nature*, 562(7726):210–216, 2018.

B. L. Elsworth, M. S. Lyon, T. Alexander, Y. Liu, P. Matthews, J. Hallett, P. Bates, T. Palmer, V. Haberland, G. D. Smith, et al. The mrc ieu opengwas data infrastructure. *bioRxiv*, 2020.

E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics*, 47(9):1091, 2015.

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, et al. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature genetics*, 49(10):1421, 2017.

A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. De Geus, D. I. Boomsma, F. A. Wright, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics*, 48(3):245–252, 2016.

G. Hemani, J. Zheng, B. Elsworth, K. H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J. Bowden, R. Langdon, et al. The mr-base platform supports systematic causal inference across the human phenome. *Elife*, 7:e34408, 2018.

International HapMap 3 Consortium and others. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311):52, 2010.

P. R. Jansen, M. Nagel, K. Watanabe, Y. Wei, J. E. Savage, C. A. de Leeuw, M. P. van den Heuvel, S. van der Sluis, and D. Posthuma. Genome-wide meta-analysis of brain volume identifies genomic loci and genes shared with intelligence. *Nature communications*, 11(1):1–12, 2020.

H. M. Kang, N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin. Efficient control of population structure in model organism association mapping. *Genetics*, 178(3):1709–1723, 2008.

S. Kelly, N. Jahanshad, A. Zalesky, P. Kochunov, I. Agartz, C. Alloza, O. A. Andreassen, C. Arango, N. Banaj, S. Bouix, C. A. Bousman, R. M. Brouwer, J. Bruggemann, J. Bustillo, W. Cahn, V. Calhoun, D. Cannon, V. Carr, S. Catts, J. Chen, J.-X. Chen, X. Chen, C. Chiapponi, K. K. Cho, V. Ciullo, A. S. Corvin, B. Crespo-Facorro, V. Cropley, P. De Rossi, C. M. Diaz-Caneja, E. W. Dickie, S. Ehrlich, F.-M. Fan, J. Faskowitz, H. Fatouros-Bergman, L. Flyckt, J. M. Ford, J.-P. Fouche, M. Fukunaga, M. Gill, D. C. Glahn, R. Gollub, E. D. Goudzwaard, H. Guo, R. E. Gur, R. C. Gur, T. P. Gurholt, R. Hashimoto, S. N. Hatton, F. A. Henskens, D. P. Hibar, I. B. Hickie, L. E. Hong, J. Horacek, F. M. Howells, H. E. Hulshoff Pol, C. L. Hyde, D. Isaev, A. Jablensky, P. R. Jansen, J. Janssen, E. G. Jönsson, L. A. Jung, R. S. Kahn, Z. Kikinis, K. Liu, P. Klauser, C. Knöchel, M. Kubicki, J. Lagopoulos, C. Langen, S. Lawrie, R. K. Lenroot, K. O. Lim, C. Lopez-Jaramillo, A. Lyall, V. Magnotta, R. C. W. Mandl, D. H. Mathalon, R. W. McCarley,

S. McCarthy-Jones, C. McDonald, S. McEwen, A. McIntosh, T. Melicher, R. I. Mesholam-Gately, P. T. Michie, B. Mowry, B. A. Mueller, D. T. Newell, P. O'Donnell, V. Oertel-Knöchel, L. Oestreich, S. A. Paciga, C. Pantelis, O. Pasternak, G. Pearlson, G. R. Pellicano, A. Pereira, J. Pineda Zapata, F. Piras, S. G. Potkin, A. Preda, P. E. Rasser, D. R. Roalf, R. Roiz, A. Roos, D. Rotenberg, T. D. Satterthwaite, P. Savadjiev, U. Schall, R. J. Scott, M. L. Seal, L. J. Seidman, C. Shannon Weickert, C. D. Whelan, M. E. Shenton, J. S. Kwon, G. Spalletta, F. Spaniel, E. Sprooten, M. Stäblein, D. J. Stein, S. Sundram, Y. Tan, S. Tan, S. Tang, H. S. Temmingh, L. T. Westlye, S. Tønnesen, D. Tordesillas-Gutierrez, N. T. Doan, J. Vaidya, N. E. M. van Haren, C. D. Vargas, D. Vecchio, D. Velakoulis, A. Voineskos, J. Q. Voyvodic, Z. Wang, P. Wan, D. Wei, T. W. Weickert, H. Whalley, T. White, T. J. Whitford, J. D. Wojcik, H. Xiang, Z. Xie, H. Yamamori, F. Yang, N. Yao, G. Zhang, J. Zhao, T. G. M. van Erp, J. Turner, P. M. Thompson, and G. Donohoe. Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA schizophrenia DTI working group. *Mol. Psychiatry*, 23(5): 1261–1269, May 2018.

K. A. Knutson, Y. Deng, and W. Pan. Implicating causal brain imaging endophenotypes in alzheimer's disease using multivariable iwas and gwas summary data. *NeuroImage*, 223:117347, 2020.

T. J. Littlejohns, J. Holliday, L. M. Gibson, S. Garratt, N. Oesingmann, F. Alfaro-Almagro, J. D. Bell, C. Boultwood, R. Collins, M. C. Conroy, et al. The uk biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature Communications*, 11(1):1–12, 2020.

Y. Liu, S. Chen, Z. Li, A. C. Morrison, E. Boerwinkle, and X. Lin. Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *The American Journal of Human Genetics*, 104(3):410–421, 2019.

S. Marek, B. Tervo-Clemmens, F. J. Calabro, D. F. Montez, B. P. Kay, A. S. Hatoum, M. R. Donohue, W. Foran, R. L. Miller, E. Feczko, O. Miranda-Dominguez, A. M. Graham, E. A. Earl, A. J. Perrone, M. Cordova, O. Doyle, L. A. Moore, G. Conan, J. Uriarte, K. Snider, A. Tam, J. Chen, D. J. Newbold, A. Zheng, N. A. Seider, A. N. Van, T. O. Laumann, W. K. Thompson, D. J. Greene, S. E. Petersen, T. E. Nichols, B. T. Yeo, D. M. Barch, H. Garavan, B. Luna, D. A. Fair, and N. U. Dosenbach. Towards reproducible brain-wide association studies. *bioRxiv*, 2020. doi: 10.1101/2020.08.21.257758.

F. Mölder, K. P. Jablonski, B. Letcher, M. B. Hall, C. H. Tomkins-Tinch, V. Sochat, J. Forster, S. Lee, S. O. Twardziok, A. Kanitz, et al. Sustainable data analysis with snakemake. *F1000Research*, 10:33, 2021.

L. J. O'Connor, A. P. Schoech, F. Hormozdiari, S. Gazal, N. Patterson, and A. L. Price. Extreme polygenicity

of complex traits is explained by negative selection. *The American Journal of Human Genetics*, 105(3): 456–476, 2019.

M. Pividori and H. K. Im. ukbrest: efficient and streamlined data access for reproducible research in large biobanks. *Bioinformatics*, 35(11):1971–1973, 2019.

J. Qian, Y. Tanigawa, W. Du, M. Aguirre, C. Chang, R. Tibshirani, M. A. Rivas, and T. Hastie. A fast and scalable framework for large-scale and ultrahigh-dimensional sparse regression with application to the uk biobank. *PLoS genetics*, 16(10):e1009141, 2020.

S. Ripke, J. T. Walters, M. C. O'Donovan, S. W. G. of the Psychiatric Genomics Consortium, et al. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*, 2020.

X. Shen, D. M. Howard, M. J. Adams, W. D. Hill, T.-K. Clarke, I. J. Deary, H. C. Whalley, and A. M. McIntosh. A phenome-wide association and mendelian randomisation study of polygenic risk for depression in uk biobank. *Nature communications*, 11(1):1–16, 2020.

A. M. Shepherd, K. R. Laurens, S. L. Matheson, V. J. Carr, and M. J. Green. Systematic meta-review and quality assessment of the structural brain alterations in schizophrenia. *Neurosci. Biobehav. Rev.*, 36(4): 1342–1356, Apr. 2012.

S. Smith, F. Alfaro-Almagro, and K. Miller. Uk biobank brain imaging documentation. `https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf`, 2020.

S. M. Smith, G. Douaud, W. Chen, T. Hanayik, F. Alfaro-Almagro, K. Sharp, and L. T. Elliott. An expanded set of genome-wide association studies of brain imaging phenotypes in uk biobank. *Nature neuroscience*, pages 1–9, 2021.

A. Taylor-Weiner, F. Aguet, N. J. Haradhvala, S. Gosai, S. Anand, J. Kim, K. Ardlie, E. M. Van Allen, and G. Getz. Scaling computational genomics to millions of individuals with gpus. *Genome biology*, 20(1):1–5, 2019.

T. G. M. van Erp, D. P. Hibar, J. M. Rasmussen, D. C. Glahn, G. D. Pearlson, O. A. Andreassen, I. Agartz, L. T. Westlye, U. K. Haukvik, A. M. Dale, I. Melle, C. B. Hartberg, O. Gruber, B. Kraemer, D. Zilles, G. Donohoe, S. Kelly, C. McDonald, D. W. Morris, D. M. Cannon, A. Corvin, M. W. J. Machielsen, L. Koenders, L. de Haan, D. J. Veltman, T. D. Satterthwaite, D. H. Wolf, R. C. Gur, R. E. Gur, S. G. Potkin, D. H. Mathalon, B. A. Mueller, A. Preda, F. Macciardi, S. Ehrlich, E. Walton, J. Hass, V. D. Calhoun, H. J. Bockholt, S. R. Sponheim, J. M. Shoemaker, N. E. M. van Haren, H. E. Hulshoff Pol, R. A. Ophoff, R. S. Kahn, R. Roiz-Santiañez, B. Crespo-Facorro, L. Wang, K. I. Alpert, E. G. Jönsson, R. Dimitrova, C. Bois, H. C. Whalley, A. M. McIntosh, S. M. Lawrie, R. Hashimoto, P. M. Thompson,

and J. A. Turner. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol. Psychiatry*, 21(4):547–553, Apr. 2016.

H. E. Wheeler, K. Aquino-Michaels, E. R. Gamazon, V. V. Trubetskoy, M. E. Dolan, R. S. Huang, N. J. Cox, and H. K. Im. Poly-omic prediction of complex traits: Omickriging. *Genetic epidemiology*, 38(5): 402–415, 2014.

J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, et al. Common snps explain a large proportion of the heritability for human height. *Nature genetics*, 42(7):565–569, 2010.

H. Zhang, T. Schneider, C. A. Wheeler-Kingshott, and D. C. Alexander. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4):1000–1016, July 2012.

B. Zhao, T. Luo, T. Li, Y. Li, J. Zhang, Y. Shan, X. Wang, L. Yang, F. Zhou, Z. Zhu, Alzheimer's Disease Neuroimaging Initiative, Pediatric Imaging, Neurocognition and Genetics, and H. Zhu. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat. Genet.*, 51(11):1637–1644, Nov. 2019.