Inferring the Genetic Influences on Psychological Traits Using MRI Connectivity Predictive Models: Demonstration with Cognition

Alexander S. Hatoum[a, b]   Andrew E. Reineberg[a]   Philip A. Kragel[c]

Tor D. Wager[d]   Naomi P. Friedman[a, c]

[a]Institute for Behavioral Genetics, University of Colorado-Boulder, Boulder, CO, USA; [b]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA; [c]Department of Psychology and Neuroscience, University of Colorado-Boulder, Boulder, CO, USA; [d]Department of Psychological and Brain Sciences, Dartmouth University, Hanover, NH, USA

Abstract
Introduction: Genetic correlations between brain and behavioral phenotypes in analyses from major genetic consortia have been weak and mostly nonsignificant. fMRI models of systems-level brain patterns may help improve our ability to link genes, brains, and behavior by identifying reliable and reproducible endophenotypes. Work using connectivity-based predictive modeling has generated brain-based proxies of behavioral and neuropsychological variables. If such models capture activity in inherited brain systems, they may offer a more powerful link between genes and behavior. ***Method:*** As a proof of concept, we develop models predicting intelligence (IQ) based on fMRI connectivity and test their effectiveness as endophenotypes. We link brain and IQ in a model development dataset of $N = 3,000$ individuals and test the genetic correlations between brain models and measured IQ in a genetic validation sample of $N = 13,092$ individuals from the UK Biobank. We compare an additive connectivity-based model to multivariate LASSO and ridge models phenotypically and genetically. We also compare these approaches to single "candidate" brain areas. ***Results:*** We found that predictive brain models were significantly phenotypically correlated with IQ and showed much stronger correlations than individual edges. Further, brain models were more heritable (h2 = 0.155–0.181) than single brain regions (h2 = 0.038–0.118) and captured about half of the genetic variance in IQ (rG = 0.422–0.576), while rGs with single brain measures were smaller and nonsignificant. For the different approaches, LASSO and ridge were similarly predictive, with slightly weaker performance of the additive model. LASSO model weights were highly theoretically interpretable and replicated known brain IQ associations. Finally, functional connectivity models trained in midlife showed genetic correlations with early life correlates of IQ, suggesting some stability in the prediction of fMRI models. ***Conclusion:*** Multisystem predictive models hold promise as imaging endophenotypes that offer complex and theoretically relevant conclusions for future imaging genetics research.

© 2022 S. Karger AG, Basel

Introduction

The vulnerability and advantages conveyed by genetics via behavior are likely due to individuals inheriting particular brain systems that support behavior. Perhaps
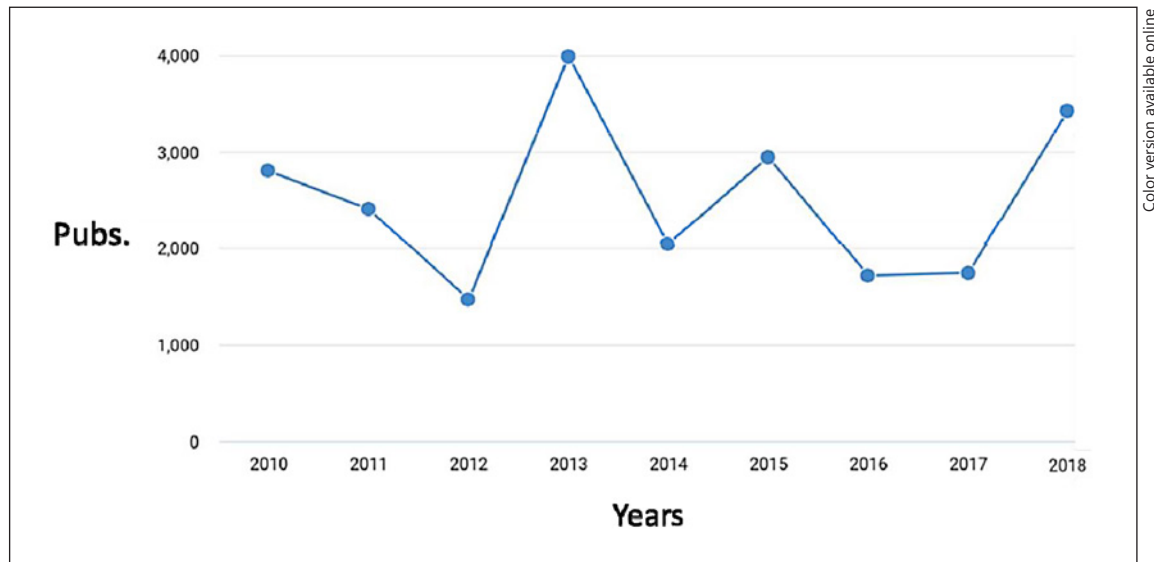
Correspondence to:
Alexander S. Hatoum, ashatoum@wustl.edu

Fig. 1. Rate of words "Imaging Genetics" and "Endophenotype" appearing together by Google Scholar search across 8 years. Pubs., publications from Google Scholar search.

this is why genetics and neuroimaging are the two most utilized biostatistical approaches for studying psychological phenomena. Luckily, these two fields are developing in parallel as both whole-genome analyses [1, 2] and multivariate brain analyses (i.e., including multiple brain areas or connections simultaneously) [3] are expanding the inferences and predictions we can make about psychological traits. As both fields expand, our ability to translate across biological psychology will depend on our ability to integrate the explosion of techniques in both fields [4].

In genetics, the standard discovery procedure is a genome-wide association study (GWAS), which requires large (10s–100s of thousands) samples and is agnostic to a priori hypothesized SNP associations. Further, these whole-genome techniques can be used to give us a measure of the genetic correlation between phenotypes or the degree to which two traits share genes and are likely inherited together.

However, the effects of individual genetic variants in GWAS are incredibly small. Given these small effect sizes and the high threshold necessary for multiple testing corrections across the genome (alpha = 5×10^{-8}), GWAS studies have necessarily adopted a coarse phenotype approach to enable larger sample sizes. We are now beginning to understand the consequences of this necessary evil as coarse phenotypes can make results difficult to interpret and can lead to less specificity in understanding

behavioral traits, as some have demonstrated with depression [5].

Further, poor phenotypes limit the types of conclusions we can make from genetic correlations and whole-genome patterns. One early suggested approach to improve genome-wide discovery and theoretical interpretation is to use intermediary traits or endophenotypes: traits that are more closely related to the genetic expression of a distal behavioral outcome [6]. That is, they should both predict behavioral traits and related genetic variations of interest in independent samples. For example, hippocampal volume is decreased in individuals with schizophrenia and non-affected related family members [7] and is likely closer to genetic inheritance than schizophrenia. Thus, it was thought that a GWAS of hippocampal volume should capture (some of) the genes influencing schizophrenia.

For years, the endophenotype approach presented an attractive alternative to the coarse approach. The most ubiquitous endophenotypes for psychological and neurological phenomena are brain and imaging genetics studies. Figure 1 shows a chart of the terms "endophenotype" and "imaging genetics" that have appeared since 2010, showing thousands of articles published on the topic each year. Further, imaging is a useful foray into the endophenotype literature as large GWAS cohorts are being assembled for MRI imaging phenotypes and offer some insight into the relative success of endophenotypes to date,

including research from the ENIGMA consortium [8, 9] and the Adolescent Brain and Cognitive Development (ABCD) Study [10].

The two goals of endophenotypes were to help discover new variants and explain the mechanisms of genetic inheritance. Unfortunately, these two goals have not yet been realized in imaging genetics. No new genetic variants influencing behavior have been discovered due to imaging endophenotypes (yet) [11]. Further, genetic associations between well-known brain endophenotypes and traits are low and nonsignificant in the largest and most powerful studies to date [12], weakening the claim that current endophenotypes are explanatory tools for scientific theory.

We argue that the imaging endophenotype approach can be improved by considering a more integrative and predictive multisystem brain approach to endophenotypes. We hypothesize that including areas from across the whole brain will capture more of the genetic variance underlying behavioral phenotypes than a single brain measure. For example, although the hippocampus has been consistently associated with schizophrenia, there was little success in finding genetic associations between hippocampal volume and schizophrenia [12]. If genes leading to the development of schizophrenia represent the inheritance of brain systems, then integrating information across the brain could increase our ability to find these genetic associations. That is, the hippocampus likely works in a system with other brain areas that more reliably predicts schizophrenia, and it is that multisystem level prediction that should be utilized in imaging genetics studies.

Multisystem Brain Predictive Models versus Candidate Approaches

While there are numerous neuroimaging brain-to-outcome associations in the literature, many of these have failed to make an impact on clinical practice. Woo, Chang, Lindquist, & Wager (2017) [13] offered a possible solution through the use of multisystem brain predictive modeling. These approaches are preferred to those focused on single candidate brain regions because they are more powerful omnibus summaries for the whole brain, allow for flexible tests of reproducibility, incorporate more information, and are more predictive than standard approaches [14]. Further, recent demonstrations of the need for larger sample sizes in imaging research have also demonstrated that whole brain approaches can reach significant and meaningful predictions in smaller samples

(of 100s of participants compared to 1,000s needed for univariate approaches [15]).

Thus, with the current approach in the literature, specific associations are found and then established as endophenotypes based on univariate tests of their association with the outcome and in related family members, akin to a "candidate endophenotype" method. It is this approach that has failed to live up to the expectations of the endophenotype claims and is not fully utilizing new trends in the imaging literature. We argue that the prediction made by these multisystem brain models can be utilized as a GWAS phenotype. In other words, we propose "multisystem brain endophenotypes" as a new path forward in the gene discovery literature. Specifically, we call the prediction made by multivariate approaches that consider many brain regions simultaneously "multisystem brain endophenotypes."

While complete literature reviews of these multisystem brain models are available elsewhere [3, 16], models of outcomes such as pain responsivity [17], sustained attention [18], Alzheimer's [19], and autism [20] have been developed and replicated across large cohorts. Vilor-Tejedor et al. [21] demonstrated a variety of multivariate models that can be used to develop brain phenotypes for genetic studies, including linear combinations like independent component analysis, principal component analysis, multivariate regression approaches, and Bayesian approaches. Interestingly, the authors also noted the wealth of consortium-level imaging data available to develop models. Their work has demonstrated that imaging is primed to develop and evaluate brain-wide endophenotypes as imaging methodologies advance [21].

Any of these models and numerous others could offer a theoretically relevant endophenotype useful for GWAS studies. Finally, while these phenotypes have statistical and theoretical advantages to standard approaches, they give the added advantage of offering many new phenotypes for GWAS discovery; for example, a GWAS of neural sustained attention patterns may inform understanding of attention-deficit/hyperactivity disorder.

Connectivity-Based Predictive Modeling

Brain imaging has many measurement approaches, or "modalities," that can be used to generate measures of the brain, for example, aspects of morphology, connectivity, or activation. Thus, a first step in developing a multisystem brain model is the choice of brain data. For our demonstration, we used functional connectivity as measured by resting-state fMRI, a method that quantifies variation in the degree to which brain regions intrinsically correlate
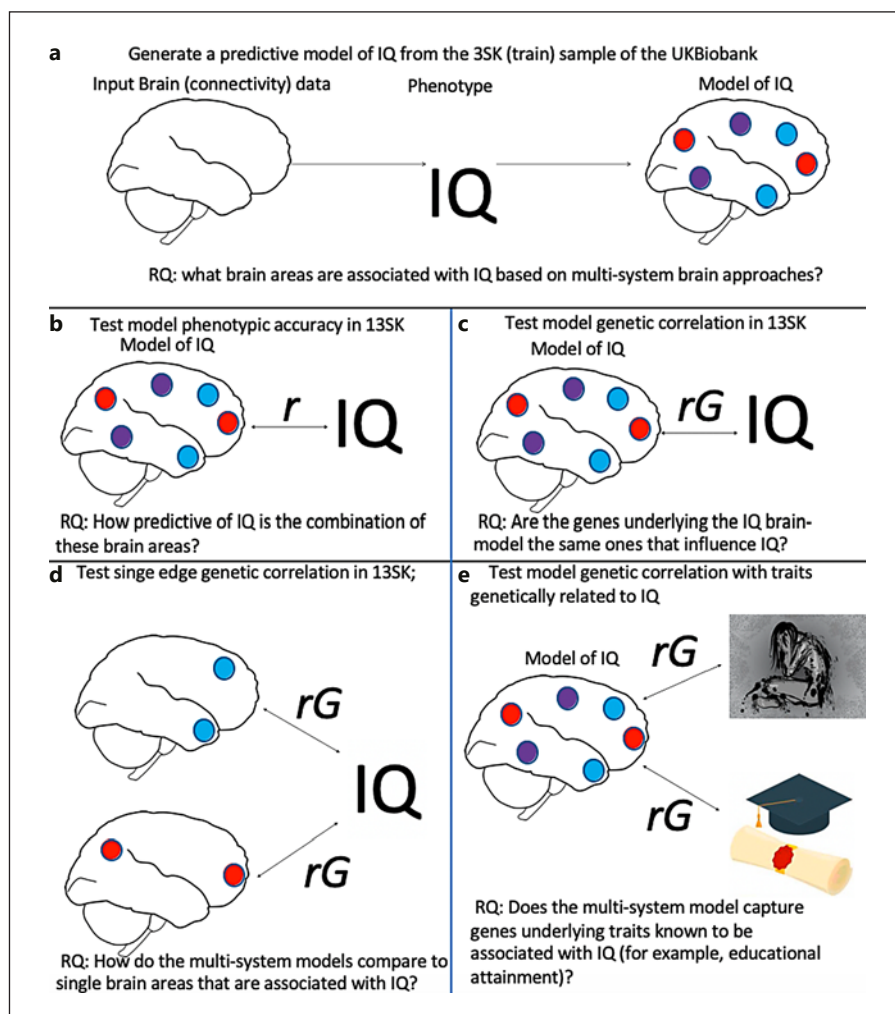
Fig. 2. Workflow for main analyses in this study. First, (**a**) we generated connectivity-based predictive models (CBPMs) of IQ in a sample of 3,000 individuals with brain data from the UK Biobank. Next, (**b**) we confirmed the out-of-sample predictive accuracy of these models phenotypically by relating brain-based predicted IQ from brain data with measured IQ in a genetic validation set of ~13,000 individuals of European descent. Because the sample of 13,000 individuals also had genetic data, (**c**) we tested the genetic correlation between predicted IQ and measured IQ in this sample. Next, (**d**) we ran the genetic correlation between the strongest single edges and IQ and compared them (using the 95% confidence intervals) to genetic correlations from multisystem brain brain-based predicted IQ. Finally, (**e**) to confirm that these genetic correlations were capturing relevant theoretical aspects of IQ, we estimated the genetic correlation between brain-based predicted IQ and known genetic correlates of IQ.

in the time course of the blood-oxygen level-dependent signal while individuals are not overtly directed toward a particular task set (i.e., are at rest) [22]. A key element of how connectivity-based predictive modeling (CBPM) is implemented using functional connectivity is the scale at which connectivity is considered – typically, all pairwise connections between 10s and 100s of regions are considered.

This Study

To demonstrate the effectiveness of multisystem brain predictive models as phenotypes for genetic research, we generate CBPMs of intelligence (IQ) and estimate the genetic correlation between brain-based predicted IQ and measured IQ. Specifically, for predictive models to be useful endophenotypes, they must be associated with the trait of interest [6] and be theoretically relevant, interpre-

table, and generalizable [11] at the genetic level. Figure 2 shows our workflow through several analyses that help achieve this aim, along with the research question answered by each analysis.

First, (a) in a sample of 3,000 individuals from the UK Biobank, we develop three predictive models of IQ with three different predictive techniques. We extract the weights from the models to see how they align with past connectivity-based studies of IQ. Next, (b–c) to test the effectiveness of the prediction of these three models as endophenotypes in a genetic validation sample of 13,092 unrelated individuals, we estimate the phenotypic and genetic correlation between predicted and measured IQ and conduct a GWAS of each model's prediction and interpretability. In this case, we use the larger sample as validation to have more power to estimate genetic correlations and more power for comparison of multisystem brain

measures and single brain measures. After estimating phenotypic and genetic correlations between brain-based predicted IQ and measured IQ, we (d) compare the multisystem brain model's predictions to the phenotypic and genetic associations between IQ and single associated edges. Finally, (e) we relate our predicted model of IQ to behaviors known to be genetically correlated with IQ.

We should note that the purpose of this work is not to test all possible machine learning methods and relevant applications (we only look at three here) but rather to create a very simple procedure for others to follow. We also report two post hoc analyses that should help develop this area of research for future genetic studies. Specifically, we use learning curves to demonstrate the improved accuracy of multisystem predictive brain models and demonstrate what sample sizes will likely be needed for this approach in the future. We then use lesion analysis to expand the theoretically relevant neurological implications of edges discovered by our machine learning approaches.

Method

Participants

The UK Biobank is a large population cohort of over 500,000 individuals [23], aged 36–72 years. We restricted our analysis to 16,092 individuals who had MRI data, IQ scores, and were of European descent. We split the sample (randomly) into a sample of 3,000 (S3K) individuals (mean age = 55.07, standard deviation age = 7.42) and 13,092 (S13K) individuals (mean age = 55.12, standard deviation age = 7.45) for different aspects of the project. Though nonintuitive, we chose to use the S13K individuals as a genetic validation set as this sample size had appropriate power to detect genetic correlations [24]. Further, the sample of 3,000 is orders of magnitude larger than most samples previously used in MRI research to develop predictive models [16, 18]. Because inference about the performance of these techniques would benefit from larger sample sizes, to estimate learning curves at increasing sample sizes (phenotypically), we use S13K to estimate the model and S3K to test the three learning techniques phenotypically in a post hoc analysis to help others establish appropriate sample sizes for training and test sets in future phenotypic work. Informed consent was obtained by the UK Biobank separately for use of cognitive ability data, MRI imaging, and genetic data (see Bycroft et al. [23] and Alfaro-Almagro et al. [25] for more).

IQ in the UK Biobank was measured at four time points: three in-person assessments and an online assessment. All assessments of IQ were 2-min tests of fluid IQ: participants had 2 min to answer as many questions as possible in a sequence of 13 questions. We estimated a latent factor model across all four time points in the full sample of ~500,000 individuals and extracted the factor score as a measure of IQ to increase reliability due to the short assessments. Data from individuals who "abandoned" the later time points were treated as missing in factor analysis. Factor scores from the imaging sample are used in this analysis.

Brain Data: Parcellation and Measurement

The UK Biobank provides highly summarized rs-fMRI data in the form of 2 full connectivity matrices per participant: pairwise correlations between 25 and 100 regions, as derived from large independent component analyses conducted with FSL's MIGP-Melodic tool. These correlations within each individual represent the training features/variables of analysis, typically called "edges." Because these 25- and 100-dimension parcellations include noise components, the actual number of signal components is lower than 25 and 100. We chose the 100-dimension parcellation, which was reduced to 55 signal components based on an analysis by the UK Biobank imaging group [25]. Thus, we had a 55×55 correlation matrix for each participant, leaving us with 1,485 unique functional connections/edges after excluding the diagonal. Each edge was residualized on age and sex before training.

SNP Data Processing and Associations

We used UK Biobank participants of European descent whose data were imputed to the Haplotype Reference Consortium [26], 1000 Genomes, and UK10K reference panels by the UK Biobank [23]. Subjects were genotyped on a UK BiLEVE array or the UK Biobank axiom array. After removing individuals with mismatched self-reported and genetic sex, we filtered imputed SNPs using a Hardy-Weinberg equilibrium p value threshold of $<1 \times 10^{-6}$, variant missingness >0.05, imputation quality score >0.95, and minor allele frequency above 0.01, retaining 7,391,068 SNPs. More information is available in a prior publication [23].

Training Procedures

We generated three scores for multisystem brain predictive models. All models were trained (with tuning for complex regression done via 10-fold cross-validation) in the training set and then applied to the nonoverlapping out-of-sample test set without additional tuning in the out-of-sample test set. First, we adapted the Shen et al. [22] CBPM procedure, which considers each edge of the connectome as an individual additive predictor. The procedure sums positively and negatively associated edges (below an associated p value threshold) and uses the negative and linear sum scores as separate beta weights in a linear model to predict the outcome of interest. We test the Shen et al. [22] model at p value cutoffs of 0.05 in an initial analysis as this is the typical value chosen in the literature [18]. In the post hoc analysis, we use learning curves to demonstrate how adjusting the p value threshold changes the prediction, if at all.

The second and third models were based on the procedure from Tobyne et al. [27] and Michel et al. [28], in which cross-validated regularized regression was used to train CBPM models. We trained ridge regression (a.k.a. L2 regularization) and Least Absolute Shrinkage and Selection Operator (LASSO) regression (a.k.a. L1 regularization) with 10-fold validation (used to choose the parameter values and regularization parameters), implemented in the R package glmnet [29]. We conducted a grid search of values from 0 to 10,000 (default of our package) for optimal regularization parameters (called "Rho" or "Lambda") for each model. To interpret the predictive models, we extracted the weights for each edge from each model. We then plotted the weights in matrix form with a set of superordinate resting-state network assignments from Yeo et al. [30] annotated along the axes to increase interpretability of the results, given high familiarity with this very popular network parcellation.

To evaluate whether certain functional connections were necessary for successful prediction, we additionally compared the LASSO and ridge regression models. This comparison evaluated whether IQ is better predicted by differences in the connectivity of a few brain regions, as characterized by LASSO regression, or is better reflected by smaller changes in the connectivity between multiple brain regions, as modeled by ridge regression. Specifically, to evaluate whether the regions identified by the LASSO model were driving the multisystem brain signal in the ridge model, the key connections identified by fitting the LASSO model were "lesioned" from the full connectivity matrix, and the remaining connections were used to fit the ridge regression model. We conducted a corrected repeated k-fold cross-validation test (10 iterations of 10-fold cross-validation; see [31]) to estimate whether LASSO or ridge regression performed better.

To see how effective these models are at predicting IQ in the training 3SK sample, we trained a ridge regression, LASSO regression, and Shen CBPM and then used the weights extracted from the 3SK sample to predict IQ in the 13SK genetic validation sample. After estimating the scores, we used a Pearson's correlation (instead of mean-squared error) to determine phenotypic association between brain-based predicted IQ (from the models' summary score) and measured IQ as we thought Pearson's r would be most comparable to the genetic correlation (rG).

Univariate Heritabilities

In the 13SK sample, we used a mixed-effects model procedure through BOLT-LMM [32] to estimate the univariate heritabilities of brain-based predicted IQ scores for the ridge regression, LASSO regression, and Shen et al. [22] procedure in the out-of-sample predictions. We also compared them to the heritability of measured IQ in the 13SK sample.

Bivariate Heritabilities

To determine if the GWAS proxies were capturing the same genes as measured IQ, we estimated the genetic correlation between measured IQ and out-of-sample brain-based predicted IQ in the 13SK genetic validation dataset. To do so, we used BOLT-LMM to estimate the genetic correlations between all three brain-based predicted IQ scores and measured IQ, separately.

To compare the genetic correlations between multisystem brain models and individual brain connections, we ran a test of association (controlling for sex and age) in the 3SK sample of each edge and IQ. After Bonferroni correction, 6 individual edges were significantly associated with IQ (Table 2) in the 3SK. We then estimated the genetic correlation between each of these 6 edges and IQ in the 13SK. We compare these to the multisystem brain correlations and correct for multiple testing in the genetic correlations (9 tests) using a Bonferroni correction.

GWAS

To see if these scores yield useful GWAS discoveries, we then used BOLT-LMM to conduct a mixed-effects GWAS (with leave-one-chromosome-out to keep from overfitting our GRM per SNP) for IQ and each predictive model in the 13SK sample. We chose BOLT-LMM to increase power and account for the polygenic background of our trait [32]. For each GWAS, we examined a Manhattan plot of individual SNP p values and also explored the observed p value deviation from expected p values using QQ plots.

Genetic Correlations with IQ Covariates

Finally, to see if brain-based predicted IQ from functional connectivity showed similar patterns of genetic correlation as IQ, we tested genetic correlations between summary statistics of the GWAS for each IQ prediction (ridge, LASSO, and Shen CBPM) and related traits through the LDhub [24, 33]. We chose 8 traits that have been previously shown to be genetically correlated with IQ in a major meta-analysis [34] and were theoretically relevant to IQ. We considered IQ proxies (educational attainment [35] and childhood IQ [36]) or psychiatric covariates of IQ (autism [37] and depressive symptoms [38]), anthropometric traits (infant head circumference [39], height [40]), and evolutionary linked traits (age of first birth [41]) as these are the typical types of hypotheses tested using genetic correlations between IQ and a covariate of IQ. We did not consider the schizophrenia and IQ genetic correlation because the schizophrenia sample that was publicly available at the time of analysis includes mixed ethnicity. We also did not run exhaustive LD score correlations between brain-based predicted IQ and all trait summary statistics available online to reduce the number of tests.

Testing Generalizability of Brain Predictive Models for IQ

To test effective sample sizes for out-of-sample prediction of IQ, we estimated learning curves for each algorithm as a function of sample size. To preview results, we chose to do this because more predictive models were also more genetically associated with measured IQ. We plotted the canonical correlation between brain-based predicted IQ and measured IQ in both the training set and test set, increasing sample sizes of 50–13,092 individuals in increments of 50. We used the S13K as the training and S3K as the validation set for this analysis to increase the spectrum of sample sizes tested. We estimate curves for the (1) ridge, (2) LASSO, (3) Shen et al. [22] at a p value threshold of 0.005, (4) p value threshold of 0.01, (5) p value threshold of 0.05, (6) p value threshold of 0.10, and (7) no p value threshold (summing up all edges). Importantly, this is the first work to compare the Shen et al. [22] predictive models at varying p value thresholds.

Results

Phenotypic Prediction

We began by training our models in the S3K sample and testing their predictive accuracy in the S3K and out-of-sample S13K. Table 1 shows the results from the phenotypic prediction in both samples, with effectiveness represented by the correlation between measured and brain-based predicted IQ. All three models showed significant phenotypic prediction in the training and test set. Phenotypic prediction in the training set ranged from $r = 0.322$–0.476. In the test set, the prediction ranged from $r = 0.187$–212. The test set may be lower because cross-validation is not a panacea for overfitting as cross-validation still capitalizes on stochastic error [42], although we employed regularization to protect against this problem. It is also likely that having a smaller training sample led to less precise cross-validated estimates of performance,
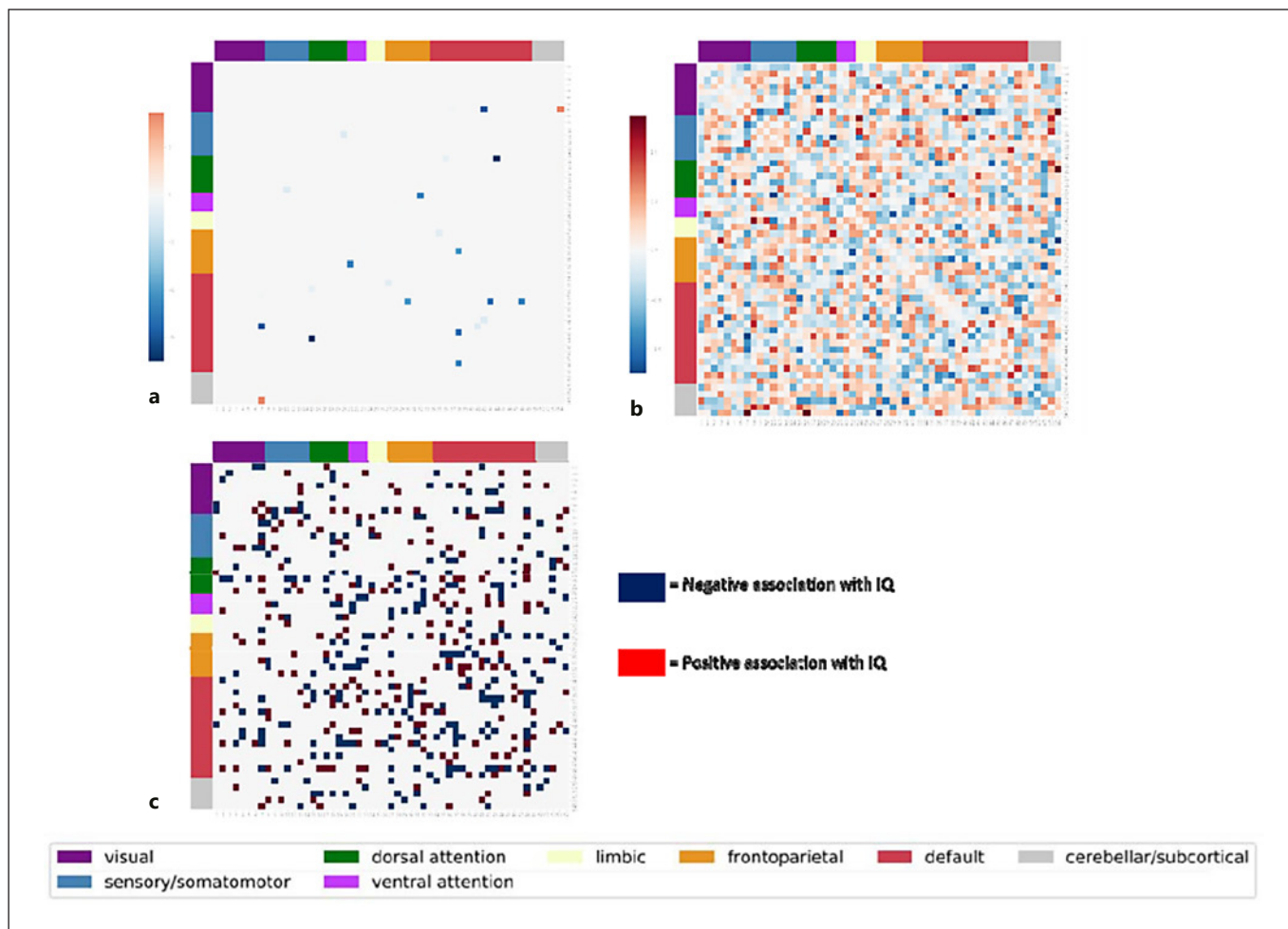
Fig. 3. Weights implicated by each predictive modeling technique organized by the Yeo 7 resting-state network to increase their interpretability. **a** Weights from the LASSO regression model. **b** Weights by the ridge regression model. **c** Weights that went into the positive (red) and negative (navy blue) sum scores for the Shen et al. procedure.

Table 1. Phenotypic prediction of IQ from each algorithm in the training and test samples

Algorithm	*r* with IQ	SE lower	SE upper	*p* value
LASSO train	0.380	0.349	0.410	<2.2e−16
Ridge train	0.476	0.447	0.503	<2.2e−16
Shen train	0.322	0.290	0.353	<2.2e−16
LASSO test	0.188	0.172	0.205	<2.2e−16
Ridge test	0.212	0.195	0.228	<2.2e−16
Shen test	0.187	0.171	0.204	<2.2e−16

Phenotypic prediction of IQ by each algorithm, measured as the Pearson's *r* between brain-based predicted IQ and measured IQ. Training was done in the 3SK sample and testing in the 13SK sample. The top three columns represent the algorithm predicting in the sample it was trained in (with LASSO and ridge using 10-fold cross-validation to guard against overfitting); the bottom three represent the out-of-sample prediction going from the 3SK to the 13SK sample.
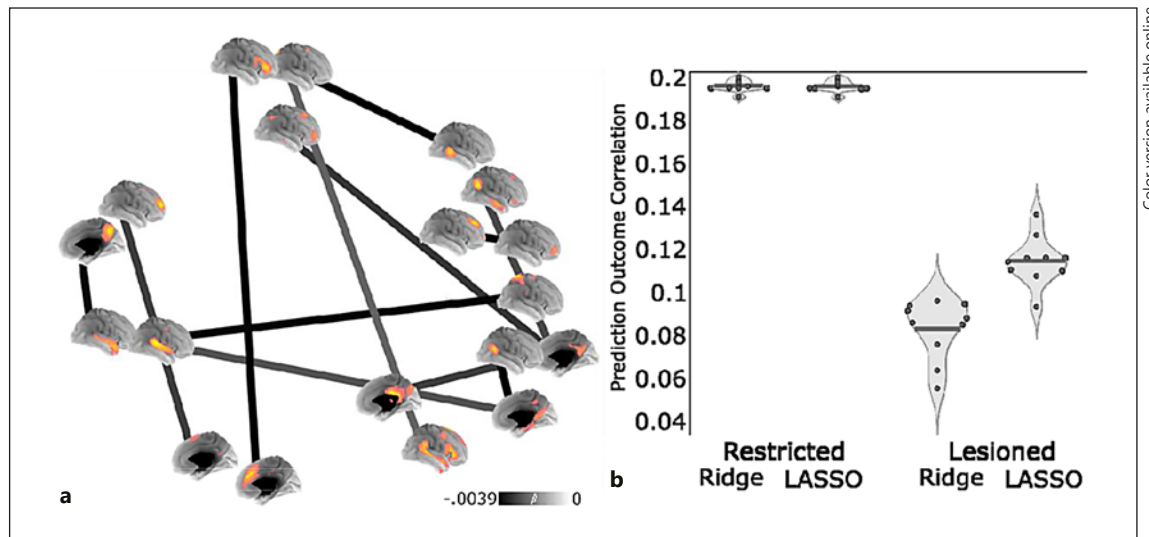
Fig. 4. Functional connections necessary for the prediction of IQ. **a** Beta estimates from the LASSO regression are depicted as a graph, where darker edges have more negative weights that indicate greater connectivity is associated with lower IQ. Surface renders depict the spatial distribution of nodes from the UK Biobank connectome, with positive loadings depicted in yellow. **b** Violin plots show the performance of models restricted to the most predictive edges identified by LASSO regression (i.e., the 12 edges shown in panel **a**) and models in which the same edges were "lesioned" or removed from the full graph. Each point in the plot corresponds to the average correlation between predicted and observed IQ for each of 10 iterations of 10-fold cross-validation. Solid lines indicate the mean across iterations.

which is a known issue for multivariate brain modeling. Ridge regression was the most predictive model in both the training and test sets (post hoc analysis shows how model predictions were affected by our sample size).

Predictive Features

For the sake of interpretability and to demonstrate the utility of this approach for theory building, we plotted edges included in the Shen et al. [22] model and weights for the ridge and LASSO regression against the Yeo 7 resting-state network parcellation in Figure 3. Due to the large sample size and number of edges, the weights from both the Shen et al. [22] procedure and the ridge regression lack a simple interpretation. In contrast, the LASSO regression was highly interpretable as only a handful of edges were needed to account for similar proportions of variance as explained by the Shen et al. [22] and the ridge models. The LASSO weights were largely negative weights on edges connecting the default mode network to other networks; that is, less connectivity between the default mode network and attention and frontal networks is predictive of higher IQ (by the LASSO output).

Because the modeling techniques produced different solutions, we hypothesized that a model using the LAS-SO-selected edges is likely the most predictive, but other edges could potentially have predictive utility. Comparisons between LASSO and ridge regression models revealed that the specific connections identified by the LAS-SO model (depicted in Fig. 4a) are driving part of the signal to predict IQ, but other edges could be used to effectively (but more weakly) predict IQ, reconciling the ridge and LASSO results. Further, a repeated k-fold cross-validation test revealed that the optimal LASSO and ridge regression models performed better when weights were estimated on LASSO-selected edges only (LASSO: mean difference = 0.1131, SD of difference = 0.0146, $t99 = 4.114$, $p < 0.0001$; ridge: mean difference = 0.0814, SD of difference = 0.0111, $t99 = 4.738$, $p < 0.0001$). These results indicate that the LASSO-selected edges informed prediction in both models, but that other connected edges are influencing the ridge and Shen et al. [22] procedures and are useful in predicting IQ.

Genetic Analyses

Univariate Heritability

First, to examine the heritability of these scores in the text set, we used BOLT-REML to run a univariate heritability analysis of each model's prediction. All three mod-

el predictions were significantly heritable (Table 2) and had similar heritability of approximately 0.15. We also tested the heritability of measured IQ in this sample, which was 0.28, higher than any predicted model.

Genetic Correlations between Predicted and Measured IQ

Next, we estimated the genetic correlation between measured IQ and the prediction of IQ by each method. All models were significantly genetically correlated with measured IQ, with moderate genetic correlations ranging from 0.422 to 0.576 (Table 2). The genetic correlations followed the same pattern as the phenotypic associations: the ridge model was most genetically correlated with measured IQ. Importantly, more than half of the genetic variance influencing IQ can be captured by ridge regression and LASSO regression brain-based predicted IQ scores. Considering that various measured IQ cohorts from the meta-analysis by Savage et al. [34] found an average rG of 0.67, our brain-based predicted IQ variable is approaching similar predictive to one of these GWAS meta-analysis cohorts.

Next, we compared genetic correlations of measured IQ with the six single edges associated with IQ in Figure 3a to genetic correlations from whole brain models. Table 2 shows the results and the 6 edges significantly phenotypically associated with IQ with Bonferroni correction. All single edges were theoretically relevant. The top associated edge was a connection involving the angular gyrus, which smaller samples have shown to be associated with IQ in the literature [43]. Several edges contained the posterior cingulate cortex, a part of the default network which is known to decrease in activation during demanding, externally directed cognitive tasks [44]; and thus, individuals with greater connectivity in the posterior cingulate may be showing less engaged task-positive network activity and lower IQ. Notably, the three brain models were nominally more heritable than the single brain edges. After correction for multiple tests, no individual edge was significantly genetically associated with measured IQ, in line with past studies reporting null effects for brain-behavior genetic correlations (9 tests). Nominally, all genetic correlations between single edges and measured IQ were lower than genetic correlations between predicted and measured IQ. This pattern is in line with our hypothesis that multisystem predictive approaches will be more useful for linking genes, brain, and behavior as they are more heritable and show higher genetic correlations than single measures of the brain.

Table 2. Heritability and co-heritability of measured IQ and the prediction of IQ by each modeling procedure

Edge/model	Network	Heritability	SE	rG with IQ	SE	rE with IQ	SE	rG CI corrected
Angular gyrus to PCC	Default to ventral attention	0.104	0.031	−0.226	0.130	−0.037	0.028	±0.361
Right frontal parietal network regions and contralateral cerebellum to PCC	Frontoparietal to ventral attention	0.070	0.030	−0.353	0.167	−0.008	0.027	±0.464
Medial superior occipital to left middle frontal and temporal structures	Dorsal attention to default mode	0.087	0.030	−0.130	0.141	−0.064	0.028	±0.390
Precuneus, lateral occipital and fusiform to middle temporal gyrus	Visual to default mode	0.102	0.030	−0.252	0.133	0.004	0.028	±0.368
Frontal pole to left frontoparietal network and cerebellum to FP	Frontoparietal to ventral attention	0.118	0.030	0.105	0.122	−0.049	0.028	±0.339
Precuneus and PCC to PCC	Default mode to default mode	0.038	0.029	−0.310	0.241	0.009	0.027	±0.669
CBPM*		0.181	0.031	0.422	0.092	0.123	0.029	±0.256
LASSO*		0.143	0.031	0.518	0.105	0.107	0.028	±0.290
Ridge*		0.155	0.031	0.576	0.100	0.116	0.028	±0.278

The heritability of each edge that was significantly associated after Bonferroni correction and the prediction of IQ and co-heritability between each edge and each predicted model of IQ with measured IQ. rG, genetic correlation between the IQ predictor and measured IQ; positive genetic correlation means that the genes that influence stronger connectivity (correlated or anti-correlated) are related to better IQ, and negative means genes that influence connectivity are related to lower IQ, rE, environmental correlation, which is the proportion of the phenotypic correlation that remains after accounting for the genetic correlation. Standard errors (SE) follow their respective statistic. We Bonferroni corrected the confidence interval to the 99.99% interval to account for multiple testing. PCC, posterior cingulate cortex. * $p < 0.05$ rG with brain-based predicted IQ post Bonferroni correction.
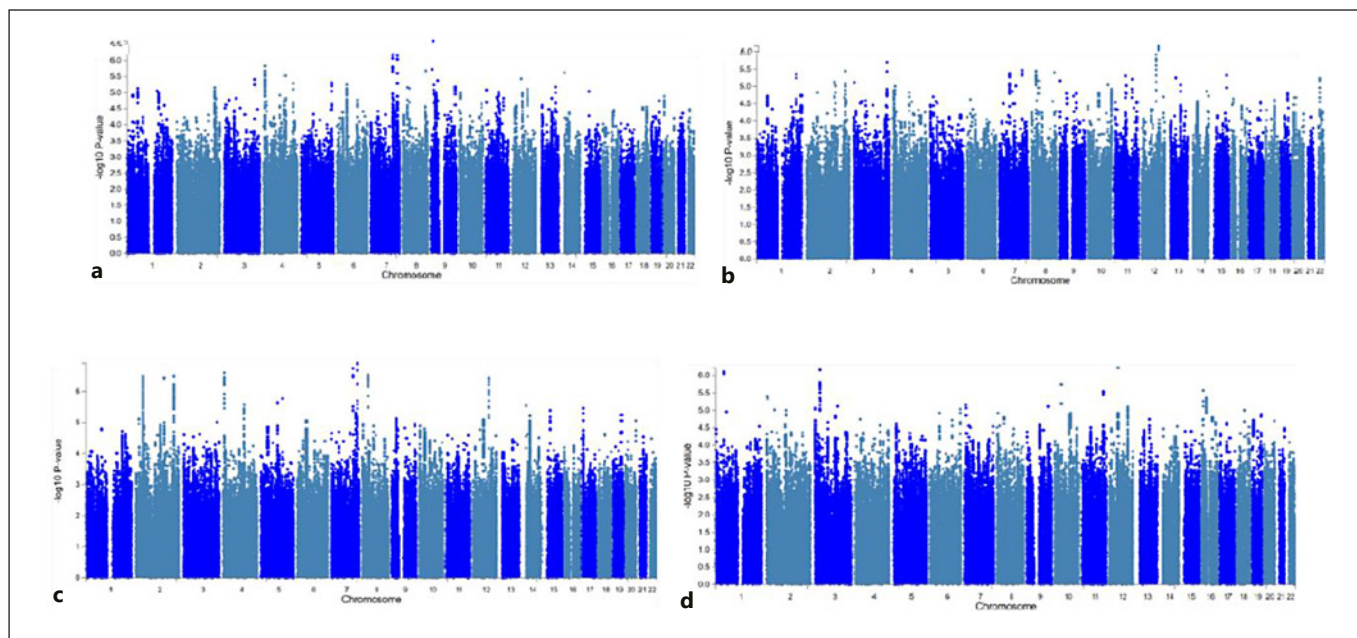
Fig. 5. Manhattan plots for genome-wide association discovery across the three predictive models (**a** is LASSO, **b** is ridge, and **c** is the Shen et al. procedure) and measured IQ (**d**) in the same individuals from the 13SK sample. p values of each SNP are organized by chromosome on the x-axis and by the −log10 of the p value on the y-axis. No p values were significant below genome-wide discovery correction ($p < 5e−8$).

GWAS of Brain-Based Predicted IQ

To test if brain-based predicted IQ is a useful GWAS proxy, we ran a GWAS with the prediction from each machine learning algorithm as the phenotype and compared it to a GWAS of measured IQ in the same individuals. No GWAS found significant SNPs above genome-wide p value significance (see Fig. 5 for Manhattan plots of each of the 4 traits). This result suggests very large GWAS samples will be needed to detect significant variants of predictive models, which is in line with sample size calculations needed for GWAS studies [45]. We used the QQ plots of estimated versus expected p values to probe the power of each of these models' predictions and measured IQ (Fig. 6). Both the LASSO and the ridge regression models had similar polygenic signals as measured IQ. However, the Shen et al. [22] procedure led to the greatest deviation from expected p values. Arguably, the Shen method may offer more power to detect variants than standard GWAS phenotypes, but this is an empirical question and did not lead to increased GWAS discovery here.

Genetic Correlations between Brain-Based Predicted IQ and Correlates of IQ

To see if brain-based predicted IQ endophenotypes were also genetically correlated with phenotypes geneti-cally correlated with measured IQ, we conducted LD score regression genetic correlations of each model predictions with 8 key traits of interest: years of education, college completion, childhood IQ, depressive symptoms, autism spectrum disorder, height, infant head circumference, and age at first birth. We compared these genetic correlations between brain-based predicted IQ and traits of interest to genetic correlations of measured IQ (in the same individuals as the predictive IQ scores were drawn from) and these same 8 key traits of interest. All three predictive models of IQ were highly genetically correlated with one another, so the genetic correlations do not change appreciably based on the method of deriving the model (ridge-LASSO rG = 0.946, CI = ±0.033; ridge-Shen rG = 0.902, CI = ±0.031; LASSO-Shen rG = 0.843, CI = ±0.044). Table 3 shows the results. IQ proxies were significantly genetically correlated with all IQ correlates except autism spectrum disorder. Interestingly, the brain-based predicted models were genetically associated with early life developmental outcomes. Notably, genetic predisposition to high IQ-predictive brain model responses was associated with a genetic predisposition to higher educational attainment (rG = 0.353 and 0.420 for years of schooling/college), larger infant head circumference (rG = 0.600), and higher childhood IQ (rG = 0.453). Thus, the
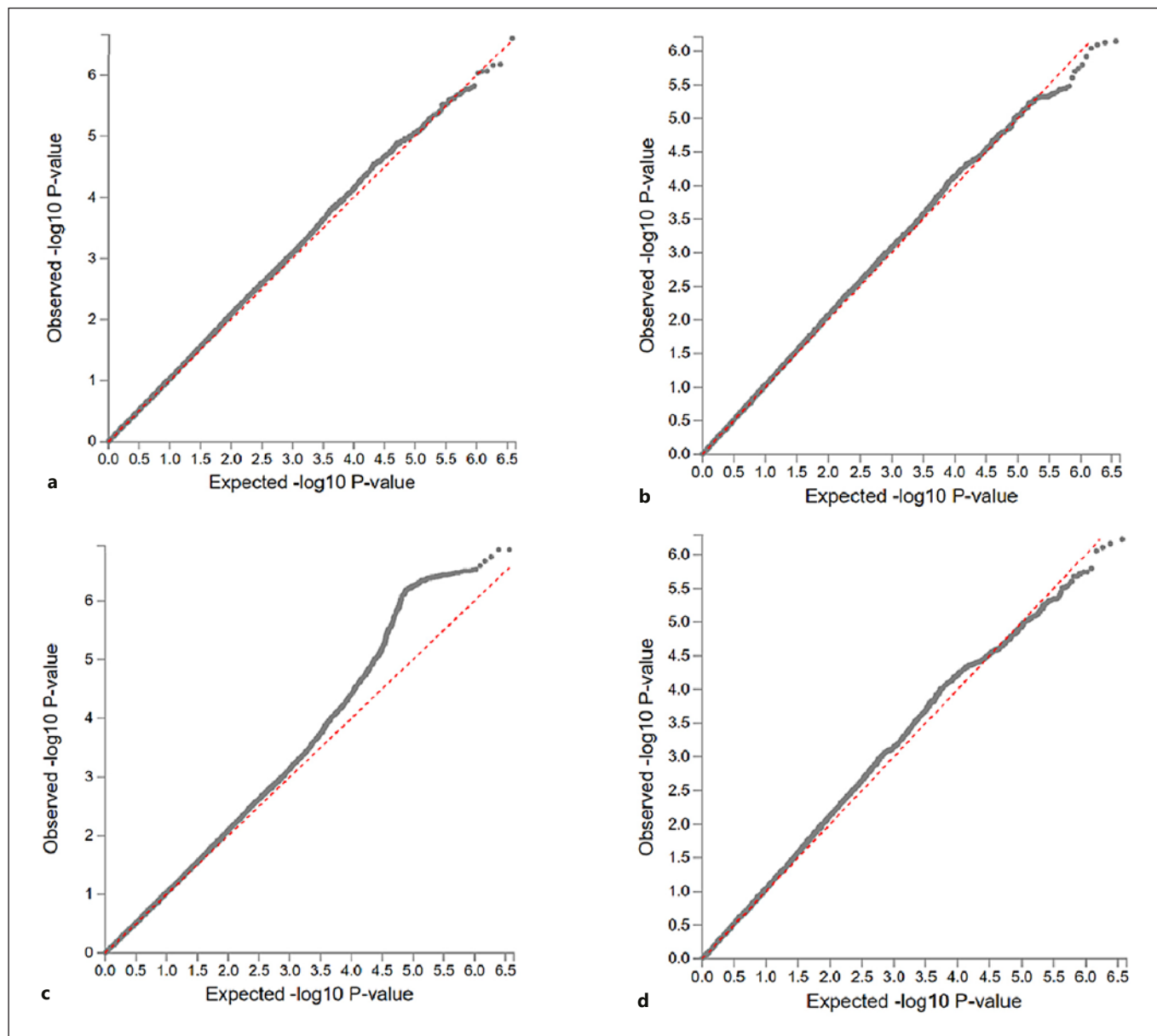
Hatoum/Reineberg/Kragel/Wager/
Friedman

Fig. 6. QQ plots for the *p* values from the genome-wide association discovery across the 3 predictive models (**a** is LASSO, **b** is ridge, and **c** is the Shen et al. procedure) and measured IQ (**d**) in the same individuals from the 13SK sample. *p* values are the dotted lines plotted based on their expected −log 10 *p* value on the *x*-axis and by their observed −log 10 *p* value on the right axis. The dashed line is the expected *p* value distribution under the null model. Deviation above the line represents signal more significant associations than expected.

brain models captured genetically linked variation shared with other IQ-related early life traits. The genes that confer a genetic advantage in early life also confer stronger IQ-related brain responses, giving some credence to the stability of the genes captured by functional connectivity in midlife.

Compared to brain-based predicted IQ, measured IQ was more strongly and significantly genetically correlated with educational attainment and was nominally more correlated with height and age at first birth to be expected from a direct (vs. proxy) measure. However, depressive symptoms and infant head circumference were nominal-

Table 3. Genetic correlations between brain-based predicted IQ and known genetic correlates of IQ

IQ model with IQ-correlated trait	RG	SE	Z	*p* value
Ridge with				
Years of schooling	0.353	0.082	4.311	<0.001
College completion	0.420	0.120	3.493	0.001
Depressive symptoms	−0.456	0.139	−3.285	0.001
Age of first birth	0.289	0.090	3.206	0.001
Childhood IQ	0.453	0.154	2.941	0.003
Infant head circumference	0.638	0.222	2.878	0.004
Height	0.023	0.062	0.370	0.711
Autism spectrum disorder	0.246	0.131	1.877	0.061
LASSO with				
Years of schooling 2016	0.381	0.113	3.359	0.001
College completion	0.397	0.148	2.687	0.007
Depressive symptoms	−0.440	0.176	−2.495	0.013
Age of first birth	0.238	0.111	2.146	0.032
Childhood IQ	0.573	0.232	2.470	0.014
Infant head circumference	0.747	0.286	2.614	0.009
Height	0.011	0.074	0.144	0.886
Autism spectrum disorder	0.263	0.170	1.550	0.121
Shen with				
Years of schooling	0.344	0.079	4.356	<0.001
College completion	0.391	0.121	3.230	0.001
Depressive symptoms	−0.428	0.148	−2.891	0.004
Age of first birth	0.248	0.095	2.613	0.009
Childhood IQ	0.298	0.159	1.871	0.061
Infant head circumference	0.676	0.242	2.793	0.005
Height	0.086	0.072	1.192	0.233
Autism spectrum disorder	0.195	0.127	1.529	0.126
Measured IQ with				
Years of schooling 2016	0.618	0.062	9.936	<0.001
College completion	0.653	0.087	7.481	<0.001
Depressive symptoms	−0.239	0.087	−2.740	0.006
Age of first birth	0.446	0.072	6.172	<0.001
Childhood IQ	0.780	0.151	5.184	<0.001
Infant head circumference	0.438	0.151	2.895	0.004
Height	0.182	0.051	3.567	<0.001
Autism spectrum disorder	0.136	0.095	1.436	0.151

Genetic correlations (rG) between brain-based predicted IQ and genetic correlates of IQ compared to genetic correlations with measured IQ. Genetic correlations were estimated using LD Score regression from summary statistics that were obtained in GWAS. Data on IQ genetic correlates are from summary statistics of major GWAS consortia.

ly more correlated with brain-based predicted IQ than measured IQ.

Utility of Predictive Models across Samples

Finally, we conducted a post hoc analysis to see what sample sizes would be needed to use in development of CBPM in the UK Biobank parcellation. Here, we also allowed the *p* value threshold of the Shen et al. [22] procedure to vary. Figure 7 shows the results of learning curves training the S13K and predicting in the S3K sample. More than 1,000 individuals are needed to reduce overfitting of the machine learning procedures (at least within this sample and parcellation). In line with results from genetic predictive models [46], the Shen et al. [22] count model seems to perform the best in smaller sample sizes (less overfitting) but becomes less useful as the sample sizes increase. The LASSO regression outperformed the ridge regression with small samples and performed worse than the ridge regression but with similar performance with larger sample sizes. Thus, in this study, LASSO regression
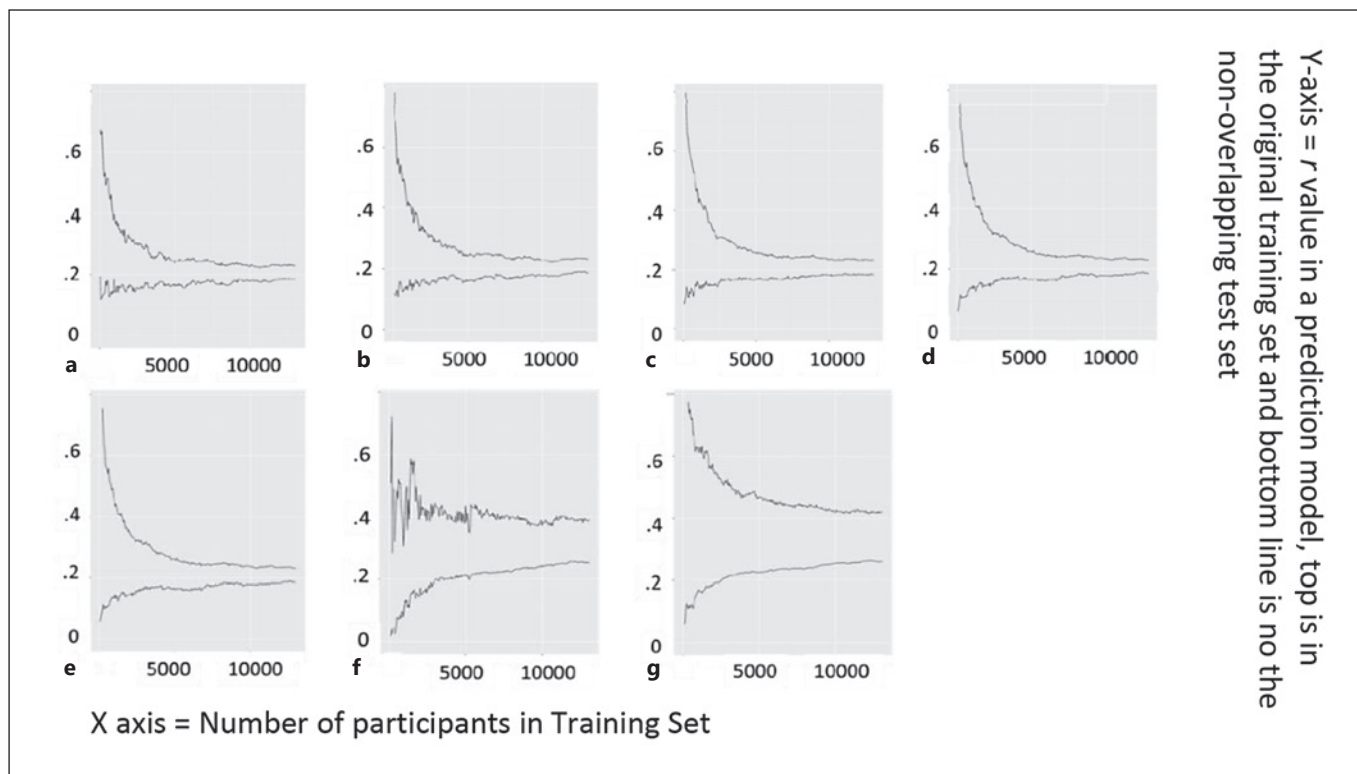
Fig. 7. Learning curves for increasing sample sizes of predictive models of IQ. The top line is the curve in the training set as sample sizes increase in the training set. The bottom line is prediction in the test set of 3,000 individuals as training sample size increasing. **a** Shen et al. *p* value threshold = 0.005. **b** Shen et al. *p* value threshold = 0.01. **c** Shen et al. *p* value threshold = 0.05, value that is often chosen. **d** Shen et al. *p* value threshold = 0.1. **e** Shen et al. *p* value threshold = 1. **f** LASSO regression. **g** Ridge regression.

offered interpretability and was a useful predictor across a range of sample sizes, meaning that LASSO is a well-balanced procedure for future work. Finally, as sample sizes became large, the ridge prediction slightly outperformed the LASSO prediction.

Discussion

We found that predictive modeling was an effective way to generate endophenotypes. We demonstrated the application of multisystem brain predictive models as GWAS proxies for discovery and theory building. We also presented novel innovations for estimation of the scores phenotypically with CBPM procedures and compared different approaches for developing these models using UK Biobank resting-state connectivity data. LASSO regression developed endophenotypes that were interpretable and predicted more than half of the genetic variance underlying IQ in out-of-sample predictions. The

Shen et al. [22] procedure seemed to offer the most power to detect polygenic signal based on deviation from expected *p* values. Conclusions for this approach are presented below.

Multisystem Brain Endophenotypes and Unmeasured Variance

We established that multisystem brain predictive models are useful endophenotypes and should be applied to work in the future. One application of this approach could be to expand the range of possible behavioral phenotypes to conduct GWAS of traits in these larger samples. For example, in the imaging literature, there are demonstrations where brain models of dichotomous traits with small samples are combined with brain models from continuously measured mechanisms to increase power, which has led to increased predictive accuracy for fibromyalgia (dichotomous) using a pain sensitivity model (continuous) [47]. This combination of brain models could expand the number of GWAS phenotypes

and direct efforts to even predict particular mechanisms underlying these traits.

When endophenotypes were first purposed as useful biomarkers, it was believed they would play a key role in gene discovery [48]. While they were largely unsuccessful at this goal, they have transitioned into useful measures for understanding underlying psychological variability [49]. As many CBPM models reflect behavioral phenotypes close to the brain and mechanisms in behavior (i.e., sustained attention in attention-deficit/hyperactivity disorder), GWAS of these models may offer a path forward in understanding the interplay of behavioral and genetic mechanisms in downstream health and wellness outcomes.

While it is often argued that predictive modeling approaches are difficult to interpret, plotting the weights can lead to insights into the brain systems underlying cognitive and behavioral phenotypes. In line with modern endophenotype approaches, we also demonstrate how prediction across the connectome is useful for theory building and align our approach with the common Yeo 7 parcellation to establish functional implications of these associations. Specifically, using LASSO regression, we found that default mode to other network connections had high utility in predicting IQ, in line with the current literature [50], and also implicated 6 edges that were associated via the standard approach of univariate tests with multiple comparison correction. We also went beyond these simple associations by discovering signal for 12 edges (with LASSO) and demonstrating that connectivity patterns beyond these 12 are associated with IQ. Thus, multisystem approaches give complementary evidence to univariate approaches and extend to more complex patterns.

Further, by comparing methods for predictions, we were able to see how the signal for predicting IQ from the connectome was (1) driven by a handful of interpretable edges and (2) related to larger multisystem brain patterns. These results are in line with imaging literature that argues IQ relies on the brain's ability to rapidly switch between different network states rather than being driven by the strength of key nodes in the network [51]. Thus, multisystem brain predictive approaches align with modern endophenotype research and brain discovery. Importantly, our approach does not limit discovery to single brain regions; more edges were weighted in the LASSO model than were associated via correction after the univariate tests of association, and the full weighted model showed much stronger genetic associations than any single edge, giving us a better picture of the multivariate pattern underlying IQ.

Further, CBPMs can be used to find genetic correlations with traits genetically related to the trait of interest. In this study, traits genetically correlated with IQ were genetically correlated with the brain-based predictions of IQ. Interestingly, infant head circumference and depressive symptoms were more genetically correlated with brain-based predicted IQ than measured IQ. This pattern may mean that the association between IQ and connectivity is strongly mediated by these phenotypes or vice versa. For example, it appears that the main edges driving the prediction of IQ (according to the LASSO output) were in the default mode network, which is thought to play a key role in depressive symptoms [52]. Therefore, it is likely that default mode connectivity genetics is influencing the overlap between IQ and depressive symptoms.

Predictive Modeling Insights from This Study
Based on patterns of results, we recommend that CBPMs utilize large samples. When comparing the different methods, the standard CBPM (Shen et al. [22]) seemed to find solutions that worked effectively in smaller samples, with adjustments to the *p* values only providing marginal change to the model performance. LASSO regression was useful across a range of sample sizes after samples exceeded ~1,000 individuals. Ridge regression seems to require very large sample sizes to avoid overfitting. The Shen sum score approach outperformed other approaches in smaller samples, but regularization in larger samples mirrored results from the statistical genetics literature on polygenic risk scores [46]. A key difference is that much smaller sample sizes are needed for the same level of prediction in multisystem brain models than in polygenic risk scores (only a couple thousand compared to 10s–100s of thousands).

There are many neuroimaging modalities across the functional and anatomical literature and many possible levels of analysis via different parcellations that can be used for multisystem brain predictive modeling. This study establishes the utility of multisystem brain modeling with the UKB functional connectivity parcellation and CBPM procedures. The procedures here predicted some, but not all, of the genetic variability underlying IQ. Though it would be outside the scope of any one manuscript to test all possible ways of predicting an outcome and its genetic associations, in the future, many other sources of data and brain parcellations may be used to expand on this method. Woo et al. [53] offer a procedure for selecting the best model that involves first testing different models in the discovery sample, selecting the one that is predictive (and makes theoretical sense), and then

testing just that model in a held-out validation dataset. Using procedures like these that have been established in imaging may improve the use of endophenotypes in the future.

Alternative Explanations for the Benefits of CBPM to Develop Endophenotypes

In our example case, endophenotypes developed from CBPMs certainly have higher genetic correlations and heritability (compared to single edges). We have broadly interpreted the genetic correlation increase in the context of a phenotypic literature showing that distributed brain systems explain individual differences in behavior. However, there are several alternative explanations for the increase in genetic correlation and heritability (compared to individual brain regions). The first is the increase in reliability. It is well known that the reliability of a phenotype is directly related to the heritability of the phenotype. When adequately powered, reweighting and including information across multiple brain edges will increase reliability and therefore heritability. However, it is worth noting that the genetic correlation is robust to the heritability inflation or deflation as the heritability is in the numerator and denominator of the equation for calculating genetic correlations (for more, see Bulik-Sullivan et al. [24]). Thus, the increases in the brain's genetic correlations cannot be explained by reliability alone.

Further, an increased genetic correlation may be due to the increase in the phenotypic overlap that is gained by "predicting" behavior. However, we should also not assume that phenotypic and genetic correlations will show the same patterns. For example, amount of tea consumed and amount of coffee consumed are phenotypically negatively correlated but show a positive genetic correlation [54].

Second, although our brain-wide endophenotypes are more heritable than the individual edges, they remain less heritable than IQ in the UK Biobank data. Indeed, most heritabilities for individual structural brain regions are also lower than our estimate for measured IQ (see Grasby et al. [47] for more). In line with lower heritability, we found few novel IQ SNPs using the brain models. Considering these results, it is worth noting that we focused only on a signal brain modality, connectivity, and our model represents a limited set of potential neurological mechanisms as such. It is likely that multimodal models (including connectivity and structure) will be needed to capture all the heritability of IQ. Indeed, it has been shown in twin studies that brain function and structure have low genetic correlations and therefore will capture different genetic variance that underlies behavior [48]. Moreover, measures of brain function appear to be less heritable than measures of brain structure [55], so multimodal models will likely be more heritable than our model based only on functional connectivity.

We have shown genuine and large gains in the genetic association using multivariate modeling approaches and different performance across the CBPM approaches used herein. It is worth noting that, though earlier endophenotypic literature claimed that closer proxies to the genetic expression would be needed, no endophenotypes to date have produced novel genetic associations (see Liu et al. [56], for a similar discussion concerning EEG measures and schizophrenia). Right now, the field of endophenotypes is limited in terms of genetic discovery. Nevertheless, we have shown that brain and behavior can have strong genetic associations with SNP data, and we are the first to find such large genetic associations. Such associations are necessary for studying the brain and behavior (e.g., [57–59]). Further development of whole brain models may reveal novel insights into the genetic architecture of behavior.

Limitations

There are a number of limitations to the study. First, our training and validation sets were closely aligned as they are both part of the UK Biobank. It is possible that these results may generalize poorly to other parcellations or subgroups (like younger cohorts) that were not used in generation of the UK Biobank parcellation. Consideration of brain parcellation and feature engineering is an important step in choosing predictive models, and this work does not argue against the importance of those considerations. Further, in using the UK Biobank parcellation, the parcellation was generated in an earlier version of the UK Biobank. We do not have the identifying information to show which individuals were used in generation of the parcellation. The use of the parcellation and the inability to identify individuals across subjects does protect individual identities but limits potential follow-up of our analysis.

In our GWAS, we focused on common variants (minor allele frequency >0.01) and highly probable variants (imputation quality score >0.95). We did so because we were running several GWAS and we wanted to conduct reasonable genome-wide screening but also reduce the computational burden. This choice means, however, that we cannot speculate about how rare variants likely influence multisystem brain models. Finally, our analyses are limited to individuals of European descent. There are not large enough samples of other ancestry groups to conduct this analysis to date, and more diverse GWAS samples are needed.

Conclusions

Multisystem brain models will be useful endophenotypes for neuropsychological outcomes in the future. Connectivity-based models performed well for predicting IQ in this study, predicting about half the genetic variance underlying IQ. Future expansions should consider other neuroimaging modalities as ways to further improve these scores.

Acknowledgment

We would like to thank Cold Spring Harbor Laboratory for hosting the preprint of this manuscript [60].

Statement of Ethics

Study approval statement and consent: this study was submitted to the University of Colorado IRB under protocol number 17-0135 but was designated as "not human subject research." Written consent was collected by the UK Biobank. Data are shared with participants, and subjects who withdrew from the study (all of whom originally gave consent) are removed before analysis.

Conflict of Interest Statement

The authors have no conflicts of interest to report.

Author Contributions

Alexander S. Hatoum conducted analysis and wrote the manuscript under the supervision of Noami P. Friedman and Tor D. Wager. Andrew E. Reineberg assisted with analysis and data visualization and provided critical feedback on the manuscript. Philip A. Kragel helped with data visualization and provided critical feedback on methodology and analysis. All authors contributed to the final written version of the manuscript.

Data Availability Statement

This research has been conducted using the UK Biobank resource under application number 24795. All data used are publicly available through the UK Biobank. Further inquiries can be directed to the corresponding author.

References

1 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

2 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011 Jan;88(1): 76–82.

3 Woo C-W, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. Nat Neurosci. 2017 Mar;20(3):365–77.

4 Hatoum AS, Reineberg AE, Smolker HR, Hewitt JK, Friedman NP. Whole-cortex mapping of common genetic influences on depression and a social deficits dimension. Transl Psychiatry. 2019 Nov;9(1):299–10.

5 Cai N, Kendler K, Flint J. Minimal phenotyping yields GWAS hits of low specificity for major depression. bioRxiv [Internet]. 2018 Oct: 440735.

6 Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry. 2003 Apr;160(4): 636–45.

7 Turetsky BI, Calkins ME, Light GA, Olincy A, Radant AD, Swerdlow NR. Neurophysiological endophenotypes of schizophrenia: the viability of selected candidate measures. Schizophr Bull. 2006 Dec;33(1):69–94.

8 Grasby KL, Jahanshad N, Painter JN, Colodro-Conde L, Bralten J, Hibar DP, et al. The genetic architecture of the human cerebral cortex. Science. 2020 Mar 20;367(6484): eaay6690.

9 Hibar DP, Westlye LT, Doan NT, Jahanshad N, Cheung JW, Ching CRK, et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA bipolar disorder working group. Mol Psychiatry. 2018;23(4):932–42.

10 Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, et al. The conception of the ABCD study: from substance use to a broad NIH collaboration. Dev Cogn Neurosci. 2018 Aug;32:4–7.

11 Hall M-H, Smoller JW. A new role for endophenotypes in the GWAS era: functional characterization of risk variants. Harv Rev Psychiatry. 2010;18(1):67–74.

12 Franke B, Stein JL, Ripke S, Anttila V, Hibar DP, van Hulzen KJE, et al. Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. Nat Neurosci. 2016 Mar;19(3):420–31.

13 Woo C-W, Krishnan A, Wager TD. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. NeuroImage. 2014 May;91:412–9.

14 Abraham A, Milham MP, Di Martino A, Craddock RC, Samaras D, Thirion B, et al. Deriving reproducible biomarkers from multisite resting-state data: an Autism-based example. NeuroImage. 2017 Feb;147:736–45.

15 Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, et al. Reproducible brain-wide association studies require thousands of individuals. Nature. 2022 Mar; 603(7902):654–60.

16 Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. NeuroImage. 2017 Jan;145(Pt B):137–65.

17 Wager TD, Atlas LY, Lindquist MA, Roy M, Woo C-W, Kross E. An fMRI-based neurologic signature of physical pain. N Engl J Med Overseas Ed. 2013 Apr;368(15):1388–97.

18 Rosenberg MD, Finn ES, Scheinost D, Papademetris X, Shen X, Constable RT, et al. A neuromarker of sustained attention from whole-brain functional connectivity. Nat Neurosci. 2016 Jan;19(1):165–71.

19 Sheng J, Wang L, Cheng H, Zhang Q, Zhou R, Shi Y. Strategies for multivariate analyses of imaging genetics study in Alzheimer's disease. Neurosci Lett. 2021 Sep;762:762136147.

20 Nogay HS, Adeli H. Machine learning [ML] for the diagnosis of autism spectrum disorder [ASD] using brain imaging. Rev Neurosci. 2020 Dec;3141(8):825–41.

21 Vilor-Tejedor N, Garrido-Martín D, Rodriguez-Fernandez B, Lamballais S, Guigó R, Gispert JD. Multivariate analysis and modelling of multiple brain endOphenotypes: let's MAMBO. Comput Struct Biotechnol J. 2021 Jan;19:5800–10.

22 Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, et al. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat Protoc. 2017 Feb;12(3):506–18.

23 Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.

24 Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015 Sep;47(11):1236–41.

25 Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and Quality Control for the first 10, 000 brain imaging datasets from UK Biobank. Neuroimage. 2018 Feb;166:400–24.

26 the Haplotype Reference Consortium, McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016 Oct;48(10):1279–83.

27 Tobyne SM, Somers DC, Brissenden JA, Michalka SW, Noyce AL, Osher DE. Prediction of individualized task activation in sensory modality-selective frontal cortex with 'connectome fingerprinting. NeuroImage. 2018 Dec;183:173–85.

28 Michel V, Gramfort A, Varoquaux G, Eger E, Thirion B. Total variation regularization for fMRI-based prediction of behavior. IEEE Trans Med Imaging. 2011 Jul;30(7):1328–40.

29 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.

30 Thomas Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 2011;106:31125–65.

31 Bouckaert RR, Frank E. Evaluating the replicability of significance tests for comparing learning algorithms. Berlin, Heidelberg: Springer; 2004. p. 3–12.

32 Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47(3):284–90.

33 Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33(2):272–9.

34 Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269, 867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018;50(7):912–9.

35 Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112–21.

36 Benyamin B, Pourcain B, Davis OS, Davies G, Hansell NK, Brion M-JA, et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. Mol Psychiatry. 2014;19(2):253–8.

37 Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51(3):431–44.

38 Okbay A, Baselmans BML, De Neve J-E, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses. Nat Genet. 2016;48(6):624–33.

39 Taal HR, St Pourcain B, Thiering E, Das S, Mook-Kanamori DO, Warrington NM, et al. Common variants at 12q15 and 12q24 are associated with infant head circumference. Nat Genet. 2012;44(5):532–8.

40 Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010;467(7317):832–8.

41 Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. Nat Genet. 2016; 48(12):1462–72.

42 Little MA, Varoquaux G, Saeb S, Lonini L, Jayaraman A, Mohr DC, et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al GigaScience. 2017;6(5):1–6.

43 Song M, Zhou Y, Li J, Liu Y, Tian L, Yu C, et al. Brain spontaneous functional connectivity and intelligence. NeuroImage. 2008;41(3):1168–76.

44 Basten U, Stelzel C, Fiebach CJ. Intelligence is differentially related to neural effort in the task-positive and the task-negative brain network. Intelligence. 2013;41(5):517–28.

45 Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012;90(1):7–24.

46 Dudbridge F. Power and predictive accuracy of polygenic risk scores. Wray NR, editor. PLoS Genet. 2013;9(3):e1003348.

47 López-Solà M, Woo C-W, Pujol J, Deus J, Harrison BJ, Monfort J, et al. Towards a neurophysiological signature for fibromyalgia. Pain. 2017;158(1):15834–47.

48 Gould TD, Gottesman II. Psychiatric endophenotypes and the development of valid animal models. Genes Brain Behav. 2006;5(2):113–9.

49 Iacono WG, Malone SM, Vrieze SI. Endophenotype best practices. Int J Psychophysiol. 2017;111:115–44.

50 Mak LE, Minuzzi L, MacQueen G, Hall G, Kennedy SH, Milev R. The default mode network in healthy individuals: a systematic review and meta-analysis. Brain Connect. 2017; 7(1):25–33.

51 Barbey AK. Network neuroscience theory of human intelligence. Trends Cogn Sci. 2018; 22(1):8–20.

52 Kaiser RH, Andrews-Hanna JR, Wager TD, Pizzagalli DA, Rj D, JJ. Large-scale network dysfunction in major depressive disorder. JAMA Psychiatry. 2015;72(6):603.

53 Woo C-W, Wager TD. Neuroimaging-based biomarker discovery and validation. Pain. 2015;156(8):1379–81.

54 Taylor AE, Davey Smith G, Munafò MR. Associations of coffee genetic risk scores with consumption of coffee, tea and other beverages in the UK Biobank. Addiction. 2018; 113(1):148–57.

55 Smith SM, Douaud G, Chen W, Hanayik T, Alfaro-Almagro F, Sharp K, et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. Nat Neurosci. 2021;24(5):737–45.

56 Liu M, Malone SM, Vaidyanathan U, Keller MC, Abecasis G, McGue M, et al. Psychophysiological endophenotypes to characterize mechanisms of known schizophrenia genetic loci. Psychol Med. 2017;47(6):1116–25.

57 Hatoum AS, Johnson EC, Baranger DAA, Paul SE, Agrawal A, Bogdan R. Polygenic Risk Scores for Alcohol Involvement Relate to Brain Structure in Substance-Naïve Children: Results from the ABCD Study. Genes Brain Behav. 2021:e12756.

58 Hatoum AS, Johnson EC, Agrawal A, Bogdan R. Brain structure and problematic alcohol use: a test of plausible causation using latent causal variable analysis. Brain Imaging Behav. 2021;15(6):2741–5.

59 Klein M, Walters RK, Demontis D, Stein JL, Hibar DP, Adams HH, et al. Genetic markers of ADHD-related variations in intracranial volume. Am J Psychiatry. 2019;176(3):228–38.

60 Hatoum AS, Reineberg AE, Kragel PA, Keller MC, Wager TD, Hewitt JK, et al. Inferring the Genetic influences on Psychological Traits Using MRI Connectivity Predictive Models: Demonstration with Cognition. bioRxiv. 2019:777821.