A method for integrating neuroimaging into genetic models of learning performance

Chintan M. Mehta[1], **Jeffrey R. Gruen**[2], and **Heping Zhang**[3]

[1]Department of Biostatistics, Yale University, 300 George Street, Suite 523, New Haven, Connecticut, 06511 (USA)

[2]Department of Pediatrics and Genetics, Yale University, 464 Congress Avenue, Suite 208, New Haven, Connecticut, 06511 (USA)

[3]Department of Biostatistics, Yale University, 300 George Street, Suite 523, New Haven, Connecticut, New Haven, Connecticut, USA

Abstract

Specific learning disorders (SLD) are an archetypal example of how clinical neuropsychological traits can differ from underlying genetic and neurobiological risk factors. Disparate environmental influences and pathologies impact learning performance assessed through cognitive exams and clinical evaluations, the primary diagnostic tools for SLD. We propose a neurobiological risk for SLD with neuroimaging biomarkers which is integrated into a genomewide association study (GWAS) of learning performance in a cohort of 479 European individuals between 8 and 21 years of age. We first identified six regions of interest (ROIs) in temporal and anterior cingulate regions where the group diagnosed with learning disability has the least overall variation, relative to the other group, in thickness, area, and volume measurements. Although we used the three imaging measures, the thickness was the leading contributor. Hence, we calculated the Euclidean distances between any two individuals based on their thickness measures in the six ROIs. Then, we defined the relative similarity of one individual according to the averaged ranking of pairwise distances from the individuals to those in the SLD group. The inverse of this relative similarity is called the neurobiological risk for the individual. Single nucleotide polymorphisms in the *AGBL1* gene on chromosome 15 had a significant association with learning performance at a genomewide level. This finding was supported in an independent cohort of 2,327 individuals of the same demographic profile. Our statistical approach for integrating genetic and neuroimaging biomarkers can be extended into studying the biological basis of other neuropsychological traits.

Corresponding Author: Heping Zhang, Department of Biostatistics, Yale University, 300 George Street, Suite 523, New Haven, Connecticut 06511, Phone: 203-785-5185, heping.zhang@yale.edu.

Code and data availability

All analysis was performed in R unless noted otherwise. Names of variables used in analysis of PING and PNC data are provided in Supplementary Table 12. Code for duplicating results in figures and tables is available upon request from authors.

Author contributions

C.M.M. developed the theory and methods, performed the analysis, created figures, and drafted the manuscript. J.R.G. contributed to the analysis and edited the manuscript. H.Z. guided the theory and methods development and data analysis, and edited the manuscript.

1 Introduction

Individuals with learning problems often suffer from diminished socioeconomic status and reduced emotional well-being (American Psychiatric Association, 2013), and treating them is an important public health priority (Castle, 2002). Five to fifteen percent of school children in the United States have a specific learning disorder (SLD) (Altarac & Saroha, 2007). The Diagnostic and Statistical Manual of Mental Disorders V (DSM-V) defines a SLD as a distinct diagnosis for cognitive deficits in perceiving and processing information that has a biological origin and that cannot be better explained by developmental, neuropsychological, or physical disorders (American Psychiatric Association, 2013). Learning disorders are, however, a latent construct manifesting along a continuous risk spectrum (Fletcher, Lyon, Fuchs, & Barnes, 2007).

The primary tools used for diagnosing a SLD are reading, writing and mathematical performance assessments. Many unrelated factors, including quality of instruction received, personal motivation, socioeconomic status, and the presence of emotional or attention disorders, also influence test performance (American Psychiatric Association, 2013; Fletcher et al., 2007). At the same time, there is considerable evidence that genetic factors influence cognitive traits related to learning performance such as reading, working memory, and episodic memory (Ando, Ono, & Wright, 2001; Donohoe, Deary, Glahn, Malhotra, & Burdick, 2013; Fletcher et al., 2007; Glahn et al., 2012; Hansell et al., 2015; Harlaar, Spinath, Dale, & Plomin, 2005; Panizzon et al., 2011).

There are many challenges in mapping genetic variants to learning performance because cognitive traits and neuropsychological (NP) traits, more generally, have a complex genetic basis (Lee et al., 2013; Okbay et al., 2016; Ripke et al., 2014). Missing heritability is endemic in genetic studies of NP traits and may be complicated by copy number variations and rare variants (Eichler et al., 2010). These issues raise considerable hurdles in genomewide association studies (GWAS) to detect single nucleotide polymorphisms (SNP) having significant relationships with NP traits (Visscher, Brown, McCarthy, & Yang, 2012).

Emerging evidence suggests that neuroimaging can offer insight into improving genomic studies of NP traits (Meyer-Lindenberg, 2012). Neuroimaging biomarkers, obtained using modalities such as magnetic resonance imaging or positron emission tomography, are heritable and have significant associations with NP traits such as depressed mood, schizophrenia, and cognitive deficits (Glahn et al., 2012; Meyer-Lindenberg & Weinberger, 2006). Since brain function has a closer biological relationship to clinical traits than genetic risk factors, neuroimaging biomarkers are suitable endophenotypes, or intermediaries, in genomic studies of cognitive and NP phenotypes (Donohoe et al., 2013; Glahn et al., 2012).

With this view in mind, several recent large-scale research initiatives performed neuroimaging scans with one goal being to improve genomic studies of NP traits. Integrating multimodal data collected by these initiatives into a single inferential framework requires statistical methods that account for large numbers of observed biomarkers and traits. Complex and multivariate relationships between genetics, neurobiology, and clinical assessments are not well characterized and thus handicap inference in this field called imaging genetics (J. Liu & Calhoun, 2014). Developing effective statistical frameworks for imaging genetics offers a promising avenue for better inferring sources of genetic and neurobiological risk affecting NP traits (Meyer-Lindenberg, 2012).

We sought to develop an integrated framework for studying the genomic basis of learning ability from neuroimaging, cognitive, and genetic data collected by the Pediatric Imaging, Neurocognition, and Genomics Study (PING) (Jernigan et al., 2015). To this end we proceed in three steps as outlined in Figure 1. First, we use principal component analysis to generate a quantitative learning performance score based on cognitive assessments in the NIH Toolbox used in the PING. In this approach, an ideal score would reflect cognitive domains relevant for perceiving and processing information, while deemphasizing others such as attention or executive function. Next, we use neuroimaging biomarkers to generate a neuroimaging risk score for learning problems that takes into account functional specialization of relevant brain regions. We developed novel, robust, and computationally efficient methods using variance-based methods to develop both of these scores. Finally, we integrate the combined learning performance and neuroimaging risks into a gene-environment interaction of learning performance that accounts for population heterogeneity.

2 MATERIALS AND METHODS

2.1 PING Cohort

The primary results are from data collected by the PING Study, which examined a cohort of 1,492 typically developing children from the United States, 3 years through 21 years of age. Subjects were genotyped with the Illumina Human660W-Quad Beadchip and underwent structural MRI scans with one of 13 devices at multiple sites (see Table S1 for scanner parameters). Subjects were also accessed with age-appropriate exams from the NIH Toolbox to evaluate neuropsychological performance across several cognitive domains (Weintraub et al., 2013). The cognitive attributes evaluated by the respective exams are listed in Table S2. Protocols for data collection, informed consent, and quality assurance are described in Jernigan, et al. (2015). Non-genetic data was obtained from PING's web portal, while genetic data was obtained from the study's principal investigators. All statistical procedures were evaluated in R unless otherwise noted. Script and custom functions for duplicating all results presented here may be obtained from the corresponding author.

Our analysis was restricted to 479 PING participants of European genetic ancestry (EGA), ages 8 years through 21 years. Restrictions were placed to reduce nonlinear effects on exam scores (Akshoomoff et al., 2014) and neuroimaging biomarkers due to age and population heterogeneity. See Table S3 and Table S4 for demographic summaries and inclusion criteria, respectively, for the sample of 479 subjects considered here.

EGA was determined by principal component analysis (PCA) of the genotype data from the PING. Participants were designated as having EGA if their scores on the first two principal components were below the extreme 1st percentile of participants with self-reported European ancestry (see Figure S1 for details). The sample was also restricted to include only the elder of any pair of participants having high genetic-relatedness, which was defined as having an identity-by-descent (IBD) value greater than 0.20. There were 97 such pairs in the entire PING cohort. EGA and IBD restrictions were placed to avoid inflated test statistics that often arise in samples that are highly stratified or contain cryptic relatedness. PCA and IBD analysis were both performed with PLINK version 1.9 (Purcell et al., 2007).

Variables for household income and highest level of parental education were used as covariates in the analysis. Missing values for either variable were imputed with the missForest package in R (Stekhoven & Bühlmann, 2012; Stekhoven, 2013) (see Table S5 for details).

2.2 Cognitive exam scores

Covariate effects of age, sex, household income, parents' education, and first two PCs of genotype were removed from scores on each cognitive exam through multivariate linear regression over the PING sample (n=479). Exam scores were then each scaled to zero mean and unit variance. A learning performance score was defined as the equal-weighted average of the List Sorting, Picture Sequence Memory, Picture Vocabulary, and Oral Reading Recognition exams. Executive function scores are defined as the equal-weighted average of the Attention, Flanker Inhibitory Control, Dimensional Change Card Sort, and Pattern Comparison Processing Speed exams.

2.3 Neuroimaging data

Common measures of brain morphometry for cortical and subcortical regions of interest (ROI), as well as white matter tracts associated with diffusion tensor imaging, were estimated from structural MRI scans with Freesurfer. We considered 198 neuroimaging biomarkers obtained from PING that we denote by { v_{ijk}^{OBS}} with $i = 1,…,479$ representing the subject, j= 1,…,66 the cortical ROI, and $k = 1,2,3$ the measurements of average thickness, total area, and volume respectively. Throughout our analysis, we worked with residuals from the multivariate linear regression models

$$v_{ijk}^{\mathrm{OBS}} = a_{jk}^{(0)} + \sum_{q=1}^{7} a_{jk}^{(q)} \cdot w_{iq} + a_{jk}^{(8)} \cdot w_{i;(k)}$$

for each j= 1, …, 66 and $k = 1,2,3$ over the PING sample (n=479), to control for eight covariates. These covariates include age, sex, handedness, the first two principal components of the genotype data, and two dummy variables used to represent the three sets of scanner device settings, given in Table S1, used in obtaining the subjects' neuroimaging biomarkers. Finally, the eighth covariance $w_{i;(k)}$ represents either the whole brain average thickness, total area, or volume for $k = 1,2,3$ respectively. We denote { v_{ijk}} to be residuals from these

regressions after being standardized to zero mean and unit variance for each j= 1, …,66 and $k = 1,2,3$.

Since there is evidence that household income and parental education account for variance in cortical surface area (Noble et al., 2015), we also covaried for them in the area measurements but found there was no material effect on our results.

2.4 Joint Variance test

Throughout, we use S^+ to represent 37 subjects in the PING sample (n=479) that provided an affirmative answer to whether they were ever diagnosed with a learning problem (LP). The remaining 442 subjects are represented by S^-.

The Joint Variance (JVAR) test is a method for assessing the importance of each ROI in discriminating the brain morphologies of S^+ (positive LP diagnosis) from S^- (no LP diagnosis) subgroups. Its main assumption is that the jth ROI is important if $\sigma^2_{jk+} < \sigma^2_{jk-}$ for every k = 1,2,3, where σ^2_{jk+} and σ^2_{jk-} respectively represent the population variances of $\{v_{ijk} : i \in S^+\}$ and $\{v_{ijk} : i \in S^-\}$. The idea behind this is that subjects in S^+ have greater intra-class similarity, and thus lower variance, than those in S^- along all three measurements of an important ROI.

We determined the importance of the jth ROI by testing the null hypothesis

$$\mathcal{H}_{0;j} \colon \min\left(\frac{\sigma^2_{j1+}}{\sigma^2_{j1-}}, \frac{\sigma^2_{j2+}}{\sigma^2_{j2-}}, \frac{\sigma^2_{j3+}}{\sigma^2_{j3-}}\right) \geq 1.$$

Under $\mathcal{H}_{0;j}$, variances of all three measurements of ROI j for individuals in S^+ are simultaneously greater than or equal to the respective variances for S^-. We evaluated its significance with the statistic

$$T_j := 1 - \prod_{k=1}^{3}(1 - p_{jk}),$$

where p_{jk} is the observed p-value from a one-sided ratio of variances F-test of $\sigma^2_{jk+} \geq \sigma^2_{jk-}$ with 36 and 441 degrees of freedom. Lower values of p_{jk} provide greater evidence that $\sigma^2_{jk+} < \sigma^2_{jk-}$ and imply lower T_j. The JVAR test thus rejects $\mathcal{H}_{0;j}$ for lower values of T_j.

The JVAR test statistic T_j serves as a proxy for the probability of the alternative to $\mathcal{H}_{0;j}$

$$T_j \approx P\left\{\min\left(\frac{\sigma^2_{j1+}}{\sigma^2_{j1-}}, \frac{\sigma^2_{j2+}}{\sigma^2_{j2-}}, \frac{\sigma^2_{j3+}}{\sigma^2_{j3-}}\right) < 1\right\}$$

under its null distribution. In fact, T_j is an unbiased estimate of it when v_{ij1}, v_{ij2}, and v_{ij3} are mutually independent. If the three variables are correlated, the approximation breaks down and is difficult to evaluate. Permutation tests were conducted to obtain the null distribution of the T_j's. First, $Q=10^6$ permuted samples were produced by randomly assigning 37 subjects to S^+ and the remaining 442 to S^-. The JVAR statistic T_j^q was evaluated from each randomly permuted sample. We defined $p_j:=Q^{-1}\sum_{q=1}^{Q}I\{T_j \leq T_j^q\}$ to be the p-value for the JVAR test of $\mathcal{H}_{0;j}$.

Our primary reason for taking the similarity approach is that it characterizes differences in brain morphology between S^+ and S^- in a multivariate setting to evaluate a notion of neurobiological risk in a way that is robust against non-linear relationships between neuroimaging variables. This requires selecting a manageable number of important neuroimaging biomarkers on which to place our focus. The JVAR test is ultimately a tool for performing this task in a way that is consistent with the similarity approach.

2.5 Neuroimaging similarity risk score

In the results section, we describe why we chose measure of thickness for six ROIs as being important. Denoting these variables by $v_{i1}^*, \ldots, v_{i6}^*$, with order being arbitrary, we measure "neuroimaging similarity" between each pair of subjects in the PING sample as being the Euclidean distances between these six variables. Similarity between subjects i and r is

$$D_{ir}:=\sqrt{\sum_{q}^{6}\left(v_{iq}^*-v_{rq}^*\right)^2}$$

with $D_{ir} = D_{ri}$. Relative similarities of subject i to others are the rankings $R_{i1}, \ldots, R_{in}$ of $D_{i1}, \ldots, D_{in}$ from least to greatest with $n=479$ the sample size. Lower values of D_{ir} and R_{ir} indicate greater similarity between subjects i and r, with $D_{ii}=0$ and $R_{ii}=1$ for every i.

The risk score we seek to evaluate for each is subject is their average relative similarity to subjects in S^+ (LP diagnosis). Clinical criteria for LP diagnoses however do not reflect neuroanatomical risk. As such, we considered average relative similarity to subjects in a subset of S^+ formed after identifying and removing neuroanatomical outliers.

Subject $i \in S^+$ was designated as an outlier by comparing their average distances to the group of individuals in S^+ and S^-, that is, the means of $\{ D_{ir} : r \in S^+\}$ and $\{D_{ir} : r \in S^- \}$ respectively. Letting μ_i^+ and μ_i^- denote the average distances, we quantified a level of "outlyingness" of each subject $i \in S^+$ through the p-value q_i from the Wilcoxon rank sum test of the null hypothesis $\mu_i^+ \leq \mu_i^-$. Lower values of q_i mean subject $i \in S^+$ has a lower level of outlyingness within S^+ because there is increasing evidence that $\mu_i^+ > \mu_i^-$, that is, their average distance to S^- is lower than their average distance to S^+.

We controlled for the number of outliers through a threshold v used to define the subsets by

$$S_\nu^\times = \{i \in S^+ : q_i > \nu\} \text{ and } S_\nu^+ = \{i \in S^+ : q_i \leq \nu\},$$

which respectively contain outliers and non-outliers in the LP diagnosed subgroup S^+. Selecting outliers reflects clinical LP diagnoses that do not necessarily represent neurobiological risk (American Psychiatric Association, 2013; Fletcher et al., 2007).

Each subject's rank-based average similarity with S_ν^+ is

$$Z_i^\nu = c \cdot \left(\sum_{r \in S_\nu^+} R_{ir} \right)^{-1}$$

where $c > 0$ is an arbitrary scaling value and can be set so that $\min_i Z_i^\nu = 1$. We term Z_i^ν as a risk score of "neuroimaging similarity with learning problem diagnoses" (NS-LP). Throughout our analysis, we set $\nu = 0.01$. It is easy to see that subject i will have a higher risk score for lower values of $\{R_{ir} : r \in S_\nu^+\}$, that is, for greater relative similarity to individuals in S_ν^+.

2.6 Genetic studies

We performed a genomewide association study (GWAS) of the learning performance scores over the PING sample (n=479). Marginal association tests were for performed for 488,200 SNPs on autosomal chromosomes having minor allele frequency and Hardy-Weinberg equilibrium p-value greater than 0.05 and 10^{-4}, respectively, in that sample. The genotype call rate was at least 0.99 for every subject.

Tests for the effect of each SNP were performed in two models of the learning performance score Y_i, which by construction was uncorrelated with age, sex, household income, and parental education. We included in both models the covariate effects of a learning problem diagnosis and the first two principal components of the genotype data that, along with an intercept, are represented by $\boldsymbol{b}^T \boldsymbol{W}_i$.

We first tested the null hypothesis that there was no genetic effect in the gene-risk interaction (G×R) model for the gth SNP, given by

$$Y_i = \boldsymbol{b}^T \boldsymbol{W}_i + \alpha_g Z_i + \beta_g X_{ig} + \gamma_g (Z_i \times X_{ig})$$

where X_{ig} is the number of risk alleles and $Z_i := Z_i^\nu$ is the NS-LP risk score for $\nu = 0.01$. There is no genetic effect in this model only if $\beta_g + \gamma_g \cdot Z_i = 0$ for all values of Z_i, which is possible only under the joint null hypothesis $\mathscr{H}_0^{(g)} : \beta_g = 0$, $\gamma_g = 0$. We performed a F-test with

2 and 472 df for each $\mathscr{J}_0^{(g)}$. For comparison we also performed a F-test, with 1 and 474 df, of the null hypothesis $\mathscr{M}_0^{(g)}:\theta_g=0$ in the main effect model $Y_i = \boldsymbol{b}^T\boldsymbol{W}_i + \theta_g X_{ig}$ for each SNP.

All estimates were obtained using ordinary least squares (OLS). In the G×R model, the estimated effect from one additional risk allele is $E(z;\beta,\gamma) = \beta + \gamma \cdot z$ with standard error being $SE(z;\beta,\gamma) = \sqrt{d_z^T \sum_D d_z}$, where $d_z = (1, z)^T$ is a vector of size two and Σ_D is the 2×2 matrix of the covariance for the β and γ estimates.

In our GWAS, we performed tests of the null hypotheses $\mathscr{M}_0^{(g)}$ and $\mathscr{J}_0^{(g)}$ for each of 488,200 SNPs. To correct for multiple testing, we evaluated false discovery rates over the Q=976,400 tests performed in the GWAS. We also evaluated permutation-based family-wide error rates (p-FWER) to account for correlations between SNPs and control for finite sample size.

The steps for evaluating p-FWER are as follows. Let $p_1^0, \ldots, p_Q^0$ denote the Bonferroni-adjusted p-values, sorted from least to greatest, observed in the GWAS with Q tests. In p-FWER, the Q tests are performed again, except after randomly permuting the responses over the sample. This was repeated for a total of R times. For each permutation $r=1, \ldots, R$, we denote $p_1^r, \ldots, p_Q^r$ as the Bonferroni-adjusted p-values, ordered from least to greatest.

$$\text{For } q=1, \ldots Q, \text{let} \widetilde{p_q} := \frac{1}{R}\sum_{r=1}^{R} I\{p_q^0 < q_q^r\}$$

denote the average number of times that p_q^0 is less than $p_q^1, \ldots, p_q^R$. Defining the integer $s_\alpha := \max\{s : \widetilde{p}_s \leq \alpha\}$ for $\alpha > 0$, the s_α tests with the lowest observed p-values are designated as having genomewide significance at level α. The p-FWER thus examines the joint distribution of extreme test statistics under null distributions.

2.7 Power analysis

In the JVAR test for ROI j, we compared the ratio of variances along the respective three measurements between S^+ (n=37) and S^- (n=442) subgroups in the PING sample. Through simulation experiments, we determined that the JVAR tests of size 0.05 have 80% power when $\max\left(\frac{\sigma_{j1+}^2}{\sigma_{j1-}^2}, \frac{\sigma_{j2+}^2}{\sigma_{j2-}^2}, \frac{\sigma_{j3+}^2}{\sigma_{j3-}^2}\right) \leq 0.85$ for case and control groups of these sizes.

The GWAS was conducted over the PING sample (n=479). We performed power analysis for minimum effect size needed for a sample of this size for tests of level 5×10^{-8} that achieve 80% power. For tests of $\mathscr{J}_0^{(g)}$, parameters β_g and γ_g in the GxR effect model for a SNP with minor allele frequency p must satisfy

$$\frac{2p(1-p)(\beta_g^2+\gamma_g^2)}{\widehat{\sigma_\varepsilon^2}} \geq 0.094,$$

where $\widehat{\sigma_\varepsilon^2}$ is the variance of residuals from the model fit using NS-LP scores that are first standardized to zero mean and unit variance. We verified this relationship through numerical experiments. For tests of $\mathscr{M}_0^{(g)}$, the parameter θ_g in the main effect model must satisfy

$$\frac{2p(1-p)\theta_g^2}{\widehat{\sigma_\varepsilon^2}} \geq 0.086.$$

2.8 PNC study

We found support for the significant genetic results from the GWAS of the PING sample with participants of 8 through 21 years of age and possessing EGA from the Philadelphia Neurodevelopmental Cohort (PNC) study (Satterthwaite et al., 2014). The PNC study genotyped participants with one of six platforms, of which the Illumina Human610-Quad or Human550 (v1 and v3) platforms included six SNPs we sought to validate. Subjects were included in the PNC sample if they were genotyped with those platforms, had no major developmental problems, and possessed EGA. The latter was determined by performing PCA on the three platforms separately. Subjects were designated as having EGA if their scores on the first two PCs were both greater than the respective extreme third percentile of subjects with self-reported European ancestry.

The PNC sample had n=2,327 subjects with valid scores on the Wide Range Achievement Test (WRAT) (Wilkinson & Robertson, 2006) that was taken to be cognitive response for assessing learning performance. The WRAT is a comprehensive IQ-achievement exam commonly used in diagnosing learning disorders. Table S3 and Table S6 respectively provide demographics and inclusion criteria for the PNC sample.

We performed a F-test with 1 and 2321 df, on the null hypothesis $\mathscr{M}_{0;\mathrm{PNC}}^{(g)}:\theta_g{=}0$ in the main effect model for the six most significant SNPs in the PING GWAS. Covariates included age, sex, two dummy variables for chip platform, and an indicator whether the subject had a reading problem. Household income and parental education data was not available on dbGAP. Neuroimaging data at the region of interest level from the PNC study was also not available on dbGAP. As a result, NS-LP scores could not be evaluated and, in turn, tests of the G×R model could not be performed.

3 RESULTS

3.1 Memory and language exams best predict learning problem diagnoses

We first sought a composite measure of learning performance, which refers throughout here to the cognitive measures associated with reading, writing, or mathematics deficits in the

DSM-V model of SLDs as opposed to others like attention or executive function. To achieve this goal, we performed principal component analysis (PCA) on scores from the eight exams in the NIH Toolbox. Because exam scores were highly associated with subjects' age, sex, level of household income, highest level of parents' education, and genetic ancestry (Akshoomoff et al., 2014) we removed their effects from each exam through multivariate regression prior to PCA in order to reduce these common sources of variation.

The exam loadings on the first PC, shown in Table S7, all had the same sign, which indicates it represents a latent factor for overall cognition, along with other confounding effects, common to all eight exams. Loadings on the second PC gave more insight into achieving our goal. Four exams evaluating episodic memory, working memory, reading, and vocabulary had negative loadings. These cognitive domains define learning performance as being distinct from executive control and attention. Indeed, there were positive loadings on the remaining four exams, which evaluated executive function, attention, and processing speed.

Our composite learning performance score, as described in the Methods section, was an equal-weighted average of language and memory exams and was, by construction, uncorrelated with age, sex, household income, highest level of parents' education, and first two PCs of genotype data. The mean and standard deviation of the learning performance scores were $\mu = 0$ and $\sigma = 0.67$ respectively.

The learning performance score discriminated between subjects with and without a learning diagnosis. As shown in Figure S2, the learning performance score distribution for S^+ (diagnosed with LP) subgroup was markedly lower than that of S^- subgroup (KS-test: $D = 0.39$, $p < 10^{-4}$). For comparison, we also derived a composite executive function score in the same way using the four exams assessing executive function, attention, and processing speed. There was no significant difference (KS-test: $D = 0.17$, $p = 0.26$) in the distributions of that composite score between S^+ and S^-, which indicates those domains are less relevant for diagnosis of learning problems.

3.2 Thickness of ROIs in temporal lobes best characterize neurobiological risk for learning problems

We next sought to quantify neurobiological risk related to learning performance using neuroimaging biomarkers. To decouple this risk from cognitive assessments, we used the subgroup S^+ (diagnosed with LP) as a surrogate measure. Indeed, Figure S2 shows there is a large overlap in the distributions of learning performance scores of S^+ and S^- (no LP diagnosis), which indicates the LP diagnosis status does not directly reflect cognitive evaluations.

We identified neuroimaging biomarkers that best characterized S^+ as being distinct from S^- using the joint variance (JVAR) test described in the Methods section. The heat map in Figure 2 illustrates p-values for the respective ROIs from the JVAR test, where significance was assessed from a null distribution generated by permutations. Lower p-values indicate the ROI better characterizes the brain morphology of S^+ relative to S^-. Most regions did not characterize S^+ well under the JVAR test over the PING sample. See Table S12 for p-values for all 66 ROIs.

Table 1 lists the ten most significant ROIs among the JVAR test results. Only the left temporal pole had a Bonferroni-adjusted p-value less than 0.05. Six ROIs with unadjusted p-values less than 0.054 were in the temporal lobes and anterior cingulate gyri on either hemisphere. The thickness component of the JVAR statistic contributes to the significance of all six ROIs, whereas area and volume components are less consistent. We quantified neurobiological risk for learning problems with the thickness of these six ROIs. We chose a cut-off for this number of ROIs because there was a marked decrease in significance for the right transverse temporal gyrus, which was the seventh most significant ROI.

3.3 NS-LP scores predict learning performance scores

There were 7 outliers among the 37 subjects having a positive LP diagnosis with respect to these variables for outlyingness threshold $v = 0.01$. Also using these six variables, we evaluated NS-LP risk scores, denoted $Z_i := Z_i^v$ for $v = 0.01$, in the PING sample (n=479).

The distribution of the NS-LP risk scores had a high positive skew ($\xi = 0.84$). Along with the covariates used for the neuroimaging variables, NS-LP scores also had insignificant correlations with either household income ($\hat{\rho} = 0.04$; p=0.37) or parental education ($\hat{\rho}$ =0.006; p=0.89).

To understand the relationship between the NS-LP and the learning performance scores Y_i described above, we fit three models to the PING sample (n=479). Model A is $Y_i = a_0 + a_Z Z_i$, model B is $Y_i = b_0 + b_L L_i$, and model C is $Y_i = c_0 + c_L L_i + c_Z Z_i$ with $L_i = 1$ for positive LP diagnosis and $L_i = 0$ otherwise. Learning performance scores are, by construction, uncorrelated with age, sex, household income, parents' education (highest level), and genetic ancestry. Summaries of the three model fits are given in Table 2.

Figure 3 shows there was a significant negative relationship between the NS-LP and the learning performance scores in model A ($\hat{z}_z = -0.36$, $p = 3.7 \times 10^{-4}$). It was not surprising that LP diagnosis status had a significant negative relationship to learning performance in Model B, given what was observed in Figure S2. It was surprising, however, that Model C fit learning performance better than Model B by 44% based on adjusted R^2. A F-test of the null hypothesis for Model B against the alternative of Model C yielded a significance of $p = 2.8 \times 10^{-3}$ over the PING sample. The significant negative marginal relationship between the two scores in Model C ($\hat{c}_z = 0.31$, $p = 2.8 \times 10^{-3}$) suggests that neuroimaging variables enabled NS-LP to predict learning performance.

In post-hoc analysis, we performed marginal association tests between learning performance Y_i and { v_{ijk}} for each respective $j = 1, \ldots, 66$ and $k = 1,2,3$ over the PING sample. Among the 198 tests, there was no p-value less than 0.006; see Table S10. This lends evidence to the hypothesis that multivariate approaches to neuroimaging are more favorable than univariate ones (J. Liu & Calhoun, 2014; Meyer-Lindenberg, 2012; Norman, Polyn, Detre, & Haxby, 2006).

3.4 Genetic results

We performed a GWAS of the learning performance scores under the G×R and main effect models for each of the 488,200 SNPs genotyped from the PING sample (n=479). The

significances from the GWAS of $\mathcal{M}_0^{(g)}$ and $\mathcal{J}_0^{(g)}$ tests are compared in the QQ-plots shown in Figure 4. The horizontal axis represents the theoretical null distribution of p-values over the G=488,200 tests, with values given by $-log_{10}\left(\frac{G-g}{G+1}\right)$ for g=1, …, G. The vertical axis corresponds to the ordered p-values, on a $-\log_{10}$ scale, observed in the GWAS of the two hypotheses. Tests of $\mathcal{J}_0^{(g)}$ detected two significant SNPs whereas those of $\mathcal{M}_0^{(g)}$ failed to detect any under theoretical asymptotic null distributions.

Among tests of $\mathcal{J}_0^{(g)}$ and $\mathcal{M}_0^{(g)}$ performed for G SNPs, the six tests having greatest significance were of $\mathcal{J}_0^{(g)}$ for SNPs all on the same linkage disequilibrium block within the *AGBL1* gene on chromosome 15. Table 3 provides their respective significances and model estimates. They all have FDR less than 0.05 over the 976,400 tests performed. We verified the significance with p-FWER.

The most significant SNP in both the G×R and main effect models of the 488,200 SNPs tests was rs11633708. In the G×R model, its p-value was less than 5×10^{-8}, a threshold commonly accepted for designating a SNP with genomewide significance in a GWAS. Table 3 also shows tests of the main effect model for these six SNPs over the PNC sample (n=2,327) were all significant with p-values all less than or equal to 7×10^{-4}. Table S11 provides minor allele frequencies, minor allele, and positions for these SNPs along with those of the 50 most significant SNPs in the GWAS under both the G×R and main effect models.

For these SNPs, tests of $\mathcal{J}_0^{(g)}$ had greater significance than tests of $\mathcal{M}_0^{(g)}$ suggesting that the gene-risk interaction model had greater information gain than the respective main effect model. Information gain here is the increase in adjusted R^2 of the fitted model (main effect or G×R) over that of the fitted base model, given by $Y_i = b^T W_i$. For these six SNPs, the gene-risk interaction model had higher information gain despite being penalized for having two extra parameters.

In Figure 5, the change in estimated effect from one additional allele of rs11633708, given everything else equal, on the learning performance score in the G×R model is plotted as a function of NS-LP scores (red line). A 95% confidence interval band is also shown where the upper and lower 2.5% percentiles are

$$CI_{Upper/Lower}(Z;\beta,\gamma)=E(z;\beta,\gamma)\pm 1.96 \cdot SE(z;\beta,\gamma)$$

with derivations of $E(z; \beta, \gamma)$ and $SE(z; \beta, \gamma)$ given in the methods section. The significance of the estimated effect is greatest for subjects in the sample with low NS-LP scores and is negligible for subjects scoring at the higher range of NS-LP, as indicated by the widening confidence interval that crosses 0.0 (blue line). In contrast, the main effect model (red line) does not account for differences in estimated genetic effects on learning performance, with respect to NS-LP risk, over the sample. Therefore, testing genetic effect in the main effect model suffers from diminished power.

To confirm these findings, we performed a gene-based analysis for the set of 150 SNPs on *AGBL1* included in our GWAS. The combined significance from the respective tests of $\mathscr{J}_0^{(g)}$ for these SNPs was 1.99×10^{-6} when using the VEGAS method (J. Z. Liu et al., 2010). The Bonferroni-corrected p-value is 0.04 if the same SNP-set analysis was conducted for 20,000 genes.

Finally, we performed post-hoc analysis to determine whether we could arrive at similar results using, for interaction variables in the GxR model, any of the neuroimaging biomarkers identified as being important with the JVAR test. To do this, we tested the genetic effect in G×R models for rs11633708 where the thickness measurement of the six most significant ROIs in the JVAR test are the interaction variables. The results are reported in Table S12 and show that the significance of genetic effect for these models was far lower than in the GxR model with the NS-LP score.

3.5 Sensitivity Analysis

We detected a significant association between SNPs in *AGBL1* and learning performance in the PING sample through the G×R model, which required evaluating NS-LP scores. These scores represent within-sample risk as they are constructed from pairwise similarities between subjects' neuroimaging measurements. This inherently heightens sample dependence of the genetic results.

We performed a sensitivity analysis to assess the stability of the NS-LP scores and their effect on the genetic results. To do this, we fit the main effect and G×R models for rs11633708 to a random subset of 407 subjects (85%) from the PING sample. The G×R model used NS-LP scores which were reevaluated over that subset. Figure S3 compares the information gain of the two models over 100 different random subsets. It shows that the G×R model for rs11633708 yielded more information gain than the main effect model over the random subsets.

4 DISCUSSION

We observed several SNPs in the *AGBL* gene that were significantly associated with learning performance at genomewide levels in our GWAS. *AGBL1* encodes an enzyme regulating protein deglutamylation in the cytosolic carboxypeptidase (CCP) family of catalysts. This enzyme and others in the *CCP* family have been observed to effect neuronal survival (Rogowski et al., 2010; Wang, Parris, Li, & Morgan, 2006). To the best of our knowledge, there are no findings linking *AGBL1* to any cognitive traits through either neurochemistry or genetic association studies.

On the other hand, *AGBL1* has been linked to schizophrenia in two separate GWAS with independent samples (Ikeda et al., 2013; Sullivan et al., 2008). Both GWAS identified SNPs in *AGBL1* as being the most significant among approximately 500,000 tested, albeit not at genomewide levels, with *p*-values less than 4×10^{-6}. Schizophrenic patients suffer from several cognitive deficits (Barch, 2005) including ones related to working memory (Forbes, Carrick, McIntosh, & Lawrie, 2009; Lett, Voineskos, Kennedy, Levine, & Daskalakis, 2014), episodic memory (Leavitt & Goldberg, 2009; Lepage, Sergerie, Pelletier, & Harvey, 2007),

and language (Bhati, 2005; Covington et al., 2005; Robbins, 2002). Taken together, these results suggest a pleiotropic effect of *AGBL1* on both schizophrenia and certain cognitive traits related to learning performance.

In the PING sample, the SNPs in *AGBL1* had significant association with learning performance at the genomewide level in the gene-risk interaction model. Within this model, the magnitude and direction of genetic effect depends on the NS-LP risk. The interaction term in the model mitigated the impact of this heterogeneity might have had on detecting genetic variants in *AGBL1*.

Indeed, many disparate environmental and non-genetic biological risk factors, in conjunction with genetic risk, influence how neuropsychological traits manifest (Hardy & Singleton, 2009). For example, early environmental stressors, substance abuse, hormones, and immunological factors are related to the onset of psychosis-spectrum disorders in individuals with a genetic predisposition (Agid et al., 1999; Ikeda et al., 2013; Kirkbride et al., 2006; Kulkarni, Hayes, & Gavrilidis, 2012). Identifying significant risk factors stands to dramatically improve power in statistical inference of NP traits (Agid et al., 1999). As our results demonstrate, neuroimaging biomarkers can be used to quantify neurobiological risk.

Neurobiological risk is, however, neither observable nor well-defined. Furthermore, many influences on brain plasticity and function, such as socioeconomic status (Noble et al., 2015), stress (McEwen, 1999), and cardiovascular fitness (Colcombe et al., 2004), among others (Ellingson, Fleming, Vergés, Bartholow, & Sher, 2014), are confounders present in neuroimaging studies. Here, we describe a new multivariate method that defines neurobiological risk in a sample by similarity in brain morphology for a given diagnosis (see Figure 6 for an illustration). This approach assumes only that individuals in a high diagnosis risk subgroup are clustered in a multivariate space of a few important neuroimaging biomarkers. Any given individual's risk is then their average similarity in that space to those in the diagnosed subgroup.

The clustering will remain persistent when projecting onto a single important variable and will disappear when projecting onto a single unimportant variable any individual's risk. The JVAR test identifies neuroimaging biomarkers that distinguish subjects at high risk for learning problems. The rank-based average similarity then allows for quantifying risk in a multivariate setting without the need to specify complicated parametric models of multivariate neuroimaging data. That this approach requires no additional assumptions about environmental exposures suggests that it is robust and potentially generalizable to other phenotypes and domains.

The similarity framework is also robust against "mislabeling" if the high-risk subgroup is initialized as being those with clinical diagnoses. Mislabeling here means that the diagnosis status of individuals may not match up with their neurobiological risk. Outliers in the diagnosed group can be identified through their neuroimaging biomarkers.

We used the JVAR test here to determine how well ROIs characterize the LP diagnosis group as being distinct from rest of the PING sample. As listed in Table 1, seven of the ten most significant regions are in the temporal lobes in either hemisphere. This is consistent with

well-established neuroimaging findings in learning performance. The temporal lobes contain networks vital to memory formation, recall and language comprehension (Brockway et al., 1998; Hoenig & Scheef, 2005; Leritz, Grande, & Bauer, 2006). Two other regions with high significance are in the anterior cingulate gyrus, which plays a key role in motivation through rewards-based learning (Shenhav, Botvinick, & Cohen, 2013).

Finally, the genetic results also depend on the learning performance phenotype. Here, we conditioned association on a composite of four memory and language assessments. This improves power for detecting genetic markers having significant association with latent factors common to all four assessments by reducing idiosyncratic noise and amplifying signal-to-noise ratio in the genetic model. Table S8 shows that rs11633708 had associations with the composite learning performance score with far greater significance than with any of the eight individual exams under both $\mathscr{J}_0^{(g)}$ and $\mathscr{M}_0^{(g)}$ hypotheses.

Our approach for integrating genetics, neuroimaging, and clinical data into a single statistical framework can be readily extended to other phenotypes for symptom severity, such as attention exams or psychotic symptom ratings, using risk scores of neuroimaging similarity to clinical diagnoses of attention or psychosis disorders. A major advantage to this framework is that, by summarizing neuroimaging risk with a single variable, it presents a massive reduction in tests for interactions between genetic and neuroimaging biomarkers.

When defined appropriately, neurobiological risk can better control for population heterogeneity in the study of any NP trait. Controlling for heterogeneity enables detecting genetic risk factors important for individuals at any given level of neurobiological risk. A direct consequence here is identifying targets for actionable personalized treatments of NP disorders, which is an important priority in current psychiatric and neuropsychological research missions (Insel & Cuthbert, 2015).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Agid O, Shapira B, Zislin J, Ritsner M, Hanin B, Murad H, … Lerer B. Environment and vulnerability to major psychiatric illness: A case control study of early parental loss in major depression, bipolar disorder and schizophrenia. Molecular Psychiatry. 1999; 4(2):163–172. [PubMed: 10208448]

Akshoomoff N, Newman E, Thompson WK, Mccabe C, Frazier JA, Gruen JR, … Kennedy DN. The NIH Toolbox Cognition Battery: Results from a Large Normative Developmental Sample (PING). Neuropsychology. 2014; 28(1):1–10. http://doi.org/10.1037/neu0000001.The. [PubMed: 24219608]

Altarac M, Saroha E. Lifetime prevalence of learning disability among US children. Pediatrics. 2007; 119(Suppl)(SUPPL 1):S77–83. http://doi.org/10.1542/peds.2006-2089L. [PubMed: 17272589]

American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). Washington, D.C: American Psychiatric Pub; 2013.

Ando JA, Ono YB, Wright MJC. Genetic structure of spatial and verbal working memory. Behavior Genetics. 2001; 31(6):615–624. [PubMed: 11838538]

Barch DM. The cognitive neuroscience of schizophrenia. Annual Review of Clinical Psychology. 2005; 1:321–53. http://doi.org/10.1146/annurev.clinpsy.1.102803.143959.

Bhati MT. The brain, language, and schizophrenia. Current Psychiatry Reports. 2005

Brockway JP, Follmer RL, Preuss LA, Prioleau CE, Burrows GS, Solsrud KA, … Howard J. Memory, simple and complex language, and the temporal lobe. Brain and Language. 1998; 61(1):1–29. http://doi.org/10.1006/brln.1997.1844. [PubMed: 9448928]

Castle, M. Learning Disabilities and Early Intervention Strategies: How To Reform the Special Education Referral and Identification Process. Subcommittee on Education Reform, House of Representatives, 107th Congress; 2002. p. 1

Colcombe SJ, Kramer AF, Erickson KI, Scalf P, McAuley E, Cohen NJ, … Elavsky S. Cardiovascular fitness, cortical plasticity, and aging. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(9):3316–21. http://doi.org/10.1073/pnas.0400266101. [PubMed: 14978288]

Covington MA, He C, Brown C, Naçi L, McClain JT, Fjordbak BS, … Brown J. Schizophrenia and the structure of language: The linguist's view. Schizophrenia Research. 2005; 77(1):85–98. http://doi.org/10.1016/j.schres.2005.01.016. [PubMed: 16005388]

Donohoe G, Deary IJ, Glahn DC, Malhotra aK, Burdick KE. Neurocognitive phenomics: examining the genetic basis of cognitive abilities. Psychological Medicine. 2013; 43(10):2027–36. http://doi.org/10.1017/S0033291712002656. [PubMed: 23194477]

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics. 2010; 11(6):446–50. http://doi.org/10.1038/nrg2809.

Ellingson JM, Fleming Ka, Vergés A, Bartholow BD, Sher KJ. Working memory as a moderator of impulsivity and alcohol involvement: testing the cognitive-motivational theory of alcohol use with prospective and working memory updating data. Addictive Behaviors. 2014; 39(11):1622–31. http://doi.org/10.1016/j.addbeh.2014.01.004. [PubMed: 24508184]

Fletcher, JM.; Lyon, GR.; Fuchs, LS.; Barnes, MA. Learning disabilities: From identification to intervention. New York: Guilford; 2007.

Forbes NF, Carrick LA, McIntosh AM, Lawrie SM. Working memory in schizophrenia: a meta-analysis. Psychological Medicine. 2009; 39(6):889–905. http://doi.org/10.1017/S0033291708004558. [PubMed: 18945379]

Glahn DC, Curran JE, Winkler AM, Carless Ma, Kent JW, Charlesworth JC, … Blangero J. High dimensional endophenotype ranking in the search for major depression risk genes. Biological Psychiatry. 2012; 71(1):6–14. http://doi.org/10.1016/j.biopsych.2011.08.022. [PubMed: 21982424]

Hansell NK, Halford GS, Andrews G, Shum DHK, Harris SE, Davies G, … Wright MJ. Genetic basis of a cognitive complexity metric. PloS One. 2015; 10(4):e0123886. http://doi.org/10.1371/journal.pone.0123886. [PubMed: 25860228]

Hardy J, Singleton A. Genomewide association studies and human disease. The New England Journal of Medicine. 2009; 360(17):1759–68. http://doi.org/10.1056/NEJMra0808700. [PubMed: 19369657]

Harlaar N, Spinath FM, Dale PS, Plomin R. Genetic influences on early word recognition abilities and disabilities: A study of 7-year-old twins. 2005; 46(4):373–384. http://doi.org/10.1111/j.1469-7610.2004.00358.x.

Hoenig K, Scheef L. Mediotemporal contributions to semantic processing: fMRI evidence from ambiguity processing during semantic context verification. Hippocampus. 2005; 15(5):597–609. http://doi.org/10.1002/hipo.20080. [PubMed: 15884095]

Ikeda M, Okahisa Y, Aleksic B, Won M, Kondo N, Naruse N, … Iwata N. Evidence for shared genetic risk between methamphetamine-induced psychosis and schizophrenia. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology. 2013; 38(10):1864–70. http://doi.org/10.1038/npp.2013.94. [PubMed: 23594818]

Insel TR, Cuthbert BN. Medicine. Brain disorders? Precisely. Science (New York, NY). 2015; 348(6234):499–500. http://doi.org/10.1126/science.aab2358.

Jernigan, TL.; Brown, TT.; Hagler, DJ.; Akshoomoff, N.; Bartsch, H.; Newman, E.; … Dale, AM. The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository; NeuroImage. 2015. p. 1-6.http://doi.org/10.1016/j.neuroimage.2015.04.057

Kirkbride JB, Fearon P, Morgan C, Dazzan P, Morgan K, Tarrant J, … Jones PB. Heterogeneity in incidence rates of schizophrenia and other psychotic syndromes: findings from the 3-center AeSOP study. Archives of General Psychiatry. 2006; 63(3):250–8. http://doi.org/10.1001/archpsyc.63.3.250. [PubMed: 16520429]

Kulkarni J, Hayes E, Gavrilidis E. Hormones and schizophrenia. Current Opinion in Psychiatry. 2012; 25(2):89–95. http://doi.org/10.1097/YCO.0b013e328350360e. [PubMed: 22249082]

Leavitt VM, Goldberg TE. Episodic memory in schizophrenia. Neuropsychology Review. 2009; 19(3):312–23. http://doi.org/10.1007/s11065-009-9107-0. [PubMed: 19639413]

Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, … Wray NR. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature Genetics. 2013; 45(9):984–94. http://doi.org/10.1038/ng.2711. [PubMed: 23933821]

Lepage M, Sergerie K, Pelletier M, Harvey PO. Episodic memory bias and the symptoms of schizophrenia. Canadian Journal of Psychiatry. 2007

Leritz EC, Grande LJ, Bauer RM. Temporal lobe epilepsy as a model to understand human memory: the distinction between explicit and implicit memory. Epilepsy & Behavior: E&B. 2006; 9(1):1–13. http://doi.org/10.1016/j.yebeh.2006.04.012.

Lett TA, Voineskos AN, Kennedy JL, Levine B, Daskalakis ZJ. Treating working memory deficits in schizophrenia: a review of the neurobiology. Biological Psychiatry. 2014; 75(5):361–70. http://doi.org/10.1016/j.biopsych.2013.07.026. [PubMed: 24011822]

Liu J, Calhoun VD. A review of multivariate analyses in imaging genetics. Frontiers in Neuroinformatics. 2014; 8(MAR):29. http://doi.org/10.3389/fninf.2014.00029. [PubMed: 24723883]

Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Hayward NK, Montgomery GW, Visscher PM, … MacGregor S. A versatile gene-based test for genome-wide association studies. American Journal of Human Genetics. 2010; 87(1):139–145. [PubMed: 20598278]

McEwen BS. Stress and hippocampal plasticity. Annual Review of Neuroscience. 1999; 22:105–22. http://doi.org/10.1146/annurev.neuro.22.1.105.

Meyer-Lindenberg A. The future of fMRI and genetics research. NeuroImage. 2012; 62(2):1286–92. http://doi.org/10.1016/j.neuroimage.2011.10.063. [PubMed: 22051224]

Meyer-Lindenberg A, Weinberger DR. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. Nature Reviews Neuroscience. 2006; 7(10):818–27. http://doi.org/10.1038/nrn1993. [PubMed: 16988657]

Noble KG, Houston SM, Brito NH, Bartsch H, Kan E, Kuperman JM, … Sowell ER. Family income, parental education and brain structure in children and adolescents. Nature Neuroscience. 2015; 18(5) http://doi.org/10.1038/nn.3983.

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 2006; 10(9):424–430. http://doi.org/S1364-6613(06)00184-7[pii]\r10.1016/j.tics.2006.07.005. [PubMed: 16899397]

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, … Benjamin DJ. Genome-wide association study identifies 74 loci associated with educational attainment. 2016; 533(7604):539–542. http://doi.org/10.1038/nature17671.

Panizzon MS, Lyons MJ, Jacobson KC, Franz CE, Grant MD, Eisen SA, Kremen WS. Genetic Architecture of Learning and Delayed Recall: A Twin Study of Episodic Memory. 2011; 25(4): 488–498. http://doi.org/10.1037/a0022569.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MaR, Bender D, … Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics. 2007; 81(3):559–75. http://doi.org/10.1086/519795. [PubMed: 17701901]

Ripke, S.; Neale, BM.; Corvin, A.; Walters, JTR.; Farh, K-H.; Holmans, Pa; … O'Donovan, MC. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014. http://doi.org/10.1038/nature13595

Robbins M. THE LANGUAGE OF SCHIZOPHRENIA AND THE WORLD OF DELUSION. The International Journal of Psychoanalysis. 2002; 83(2):383–405. http://doi.org/10.1516/0020757021601964.

Rogowski K, van Dijk J, Magiera MM, Bosc C, Deloulme JC, Bosson A, … Janke C. A family of protein-deglutamylating enzymes associated with neurodegeneration. Cell. 2010; 143(4):564–78. http://doi.org/10.1016/j.cell.2010.10.014. [PubMed: 21074048]

Satterthwaite TD, Elliott Ma, Ruparel K, Loughead J, Prabhakaran K, Calkins ME, … Gur RE. Neuroimaging of the Philadelphia neurodevelopmental cohort. NeuroImage. 2014; 86:544–53. http://doi.org/10.1016/j.neuroimage.2013.07.064. [PubMed: 23921101]

Shenhav A, Botvinick MM, Cohen JD. The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron. 2013; 79(2):217–40. http://doi.org/10.1016/j.neuron.2013.07.007. [PubMed: 23889930]

Stekhoven DJ. missForest: Nonparametric Missing Value Imputation using Random Forest. R Package version 1.4. 2013

Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012; 28(1):112–118. [PubMed: 22039212]

Sullivan PF, Lin D, Tzeng JY, van den Oord EJCG, Perkins D, Stroup TS, … Liu W. Genomewide association for schizophrenia in the CATIE study: results of stage 1. Molecular Psychiatry. 2008; 13(6):570–584. [PubMed: 18347602]

Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. American Journal of Human Genetics. 2012; 90(1):7–24. http://doi.org/10.1016/j.ajhg.2011.11.029. [PubMed: 22243964]

Wang T, Parris J, Li L, Morgan JI. The carboxypeptidase-like substrate-binding site in Nna1 is essential for the rescue of the Purkinje cell degeneration (pcd) phenotype. Molecular and Cellular Neurosciences. 2006; 33(2):200–13. http://doi.org/10.1016/j.mcn.2006.07.009. [PubMed: 16952463]

Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, … Gershon RC. Cognition assessment using the NIH Toolbox. Neurology. 2013; 80(11 Suppl 3)

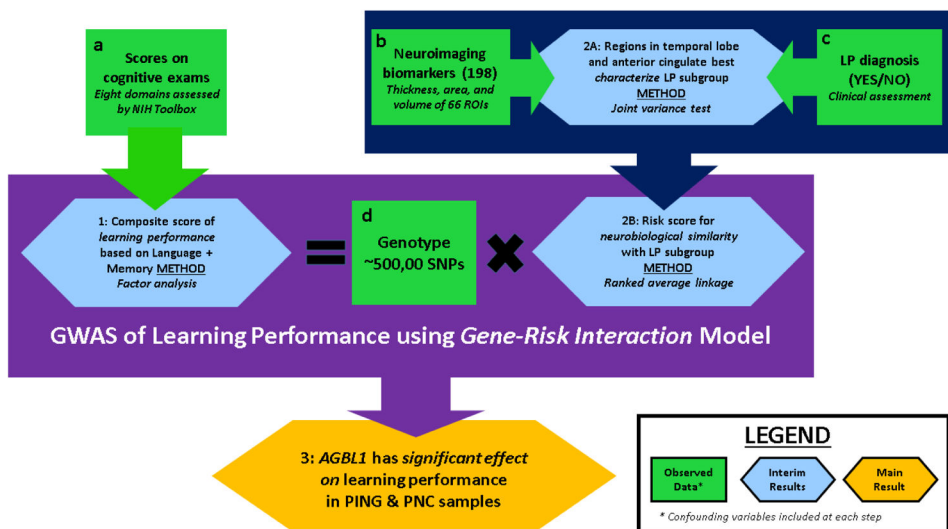Wilkinson G, Robertson G. Wide Range Achievement Test (WRAT4). 2006

Figure 1. Overview of inferential procedure applied to PING study

Data: Cognitive assessments were taken from eight exams within the NIH Toolbox. Learning problem (LP) status is a binary variable corresponding to a yes/no answer in response to the question whether a subject was ever diagnosed with a learning problem. There are 198 neuroimaging biomarkers corresponding to average thickness, area, and volume measurements for 66 cortical regions of interest (ROI) obtained from structural MRI scans. Approximately 500,000 SNPs were genotyped for each subject in the sample. Results: (1) A composite learning performance score was derived from exams assessing memory (working and episodic) and language (reading and vocabulary). (2A) The joint variance test was performed to determine how well each ROI discriminated subjects with a LP diagnosis from others in the sample using thickness, area, and volume measurements. Of the 10 most significant regions, seven were in the temporal lobe and two were in the anterior cingulate on either the left or right hemispheres. (2B) With neuroimaging biomarkers best discriminating the LP subgroup, rank-based average linkage (similarity) clustering was applied to evaluate continuous risk scores for each subjects' neuroimaging similarity to those in the LP subgroup. The similarities were with respect to the neuroimaging biomarkers identified as being most significant under the joint variance test. (3) A GWAS of the learning performance score was conducted using models including effects from interactions between genomic markers and the neuroimaging risk score. Six SNPs on the *AGBL1* gene were observed to have significant associations with learning performance score at a genomewide level.
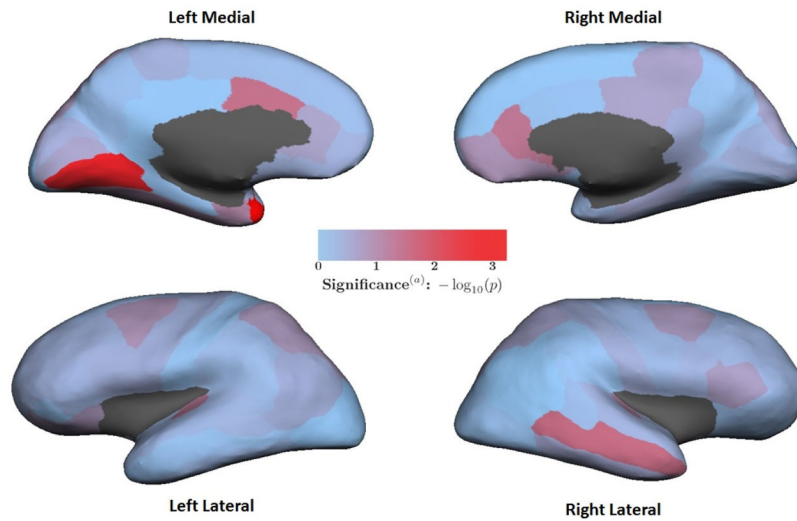
Figure 2. Heat map of each ROI's significance under JVAR test

Colors represent *p*-values from the JVAR test for each ROI comparing subjects with a learning problem diagnosis (n=37) and without a LP diagnosis (n=442) in the PING sample (n=479). Red indicates greater significance. (a) The *p*-values are not adjusted for multiple comparisons. The *p*-values from the JVAR tests for all 66 ROIs are listed in Table S9.
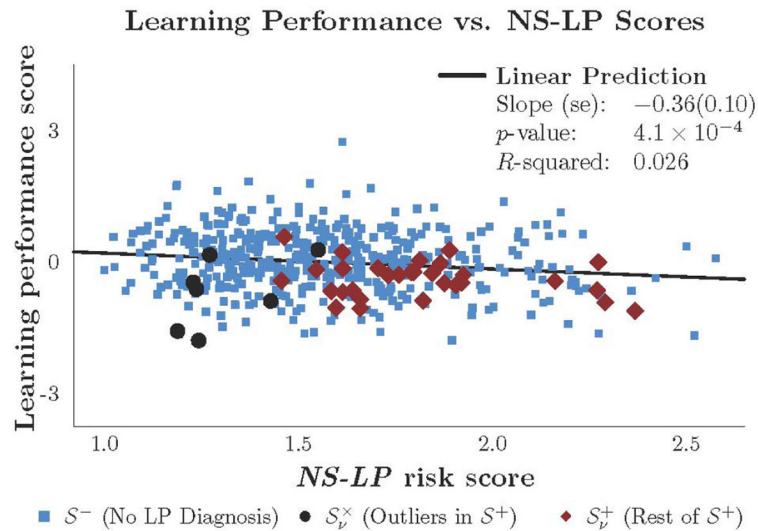
Figure 3. Learning performance scores versus NS-LP scores

The blue, black, and red markers represent subjects in subgroups denoted by S^- (n=442), $S_\nu^\times$ (n=7), and S_ν^+ (n=30) respectively. The $S_\nu^\times$ and S_ν^+ subgroups respectively contain outliers and non-outliers in S^+ (n=37) based on outlyingness of subjects, with respect to neuroanatomical measurements, threshold $\nu = 0.01$. Methods section describes how outliers are identified. Regression results are over the PING sample (n=479) and based on Model (A) specified in Table 2.
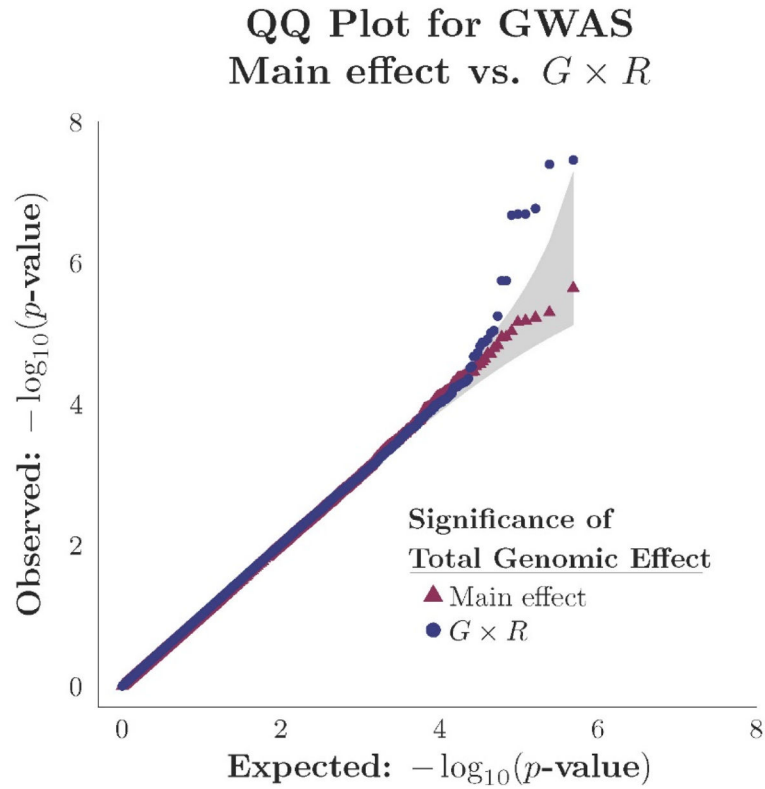
Figure 4. QQ-plot of GWAS of learning performance score in PING sample (n=479)
Tests were performed for genomic effects $G{\times}R$ and Main effect models of learning performance scores for each of 488,200 SNPs. Observed *p*-values in the GWAS of each model were sorted and then plotted on a $-log_{10}$ scale against expected values in theoretical null distributions. The shaded region indicates a 95% confidence interval. Blue dots correspond to tests of $\mathscr{J}_0^{(g)}$ in the $G{\times}R$ model. Red triangles correspond to tests of $\mathscr{M}_0^{(g)}$ in the main effect model. The genomic inflation factors for the GWAS of the $G{\times}R$ and Main effect models were 1.012 and 1.002, respectively.
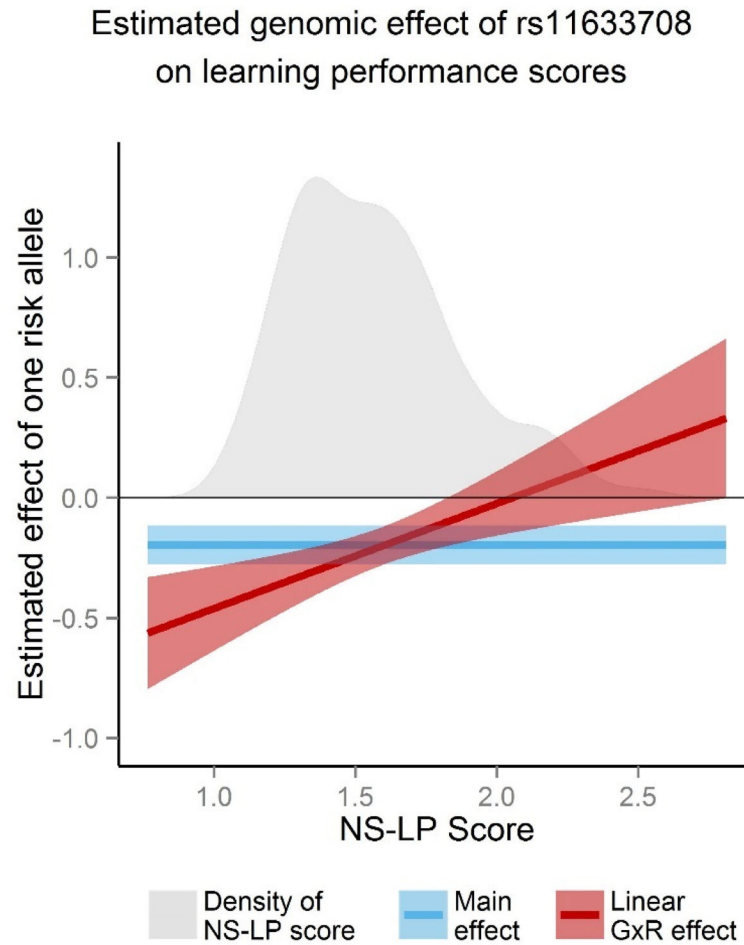
Figure 5. Estimated effect of rs11633708 on learning performance as function of NS-LP risk score

The SNP rs11633708 on *AGBL1* (chromosome 15) was the most significant in a GWAS of learning performance scores over the PING sample (n=479). The vertical axes represents estimated effect one additional risk allele of rs11633708 on learning performance scores (μ= 0, σ= 0.67). Lines and shaded regions represent estimated effects and their 95% confidence intervals in three models. Blue represents the main effect model, which does not rely on NS-LP risk scores. Red represents the *G×R* model presented in the genomic results section. Estimates and standard errors in the *G×R* model are a function of NS-LP scores. Standard errors there were evaluated from the covariance matrix of estimates in the respective linear regression models. The gray region in both plots is the empirical density of NS-LP scores and is at scale with the vertical axis.
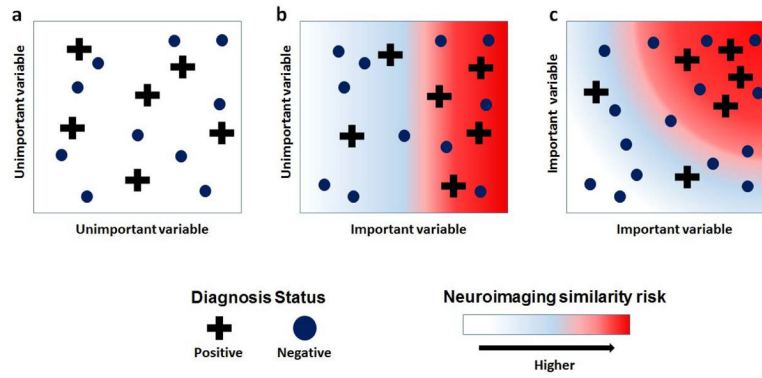
Figure 6. Illustration of neuroimaging similarity for a diagnosis

Two subgroups of subjects are formed on the basis of observed diagnosis. Plus signs represent S^+ (positive diagnosis) and dots represent S^- (negative diagnosis). Along important variables, S^+ has greater level of within-group similarity (homogeneity) relative to S^-. Along unimportant variables, the level of within-group similarity for S^+ and S^- are roughly equal. Background colors represent subjects' continuous level of along respective axes. Each figure represents bivariate comparisons between variables. Background color represent continuous level to neuroimaging similarity with S^+ in bivariate setting. (a) Axes represent two unimportant variables where neuroimaging similarity risk cannot be detected. (b) Respective axes are important and unimportant variables. Neuroimaging similarity risk increases only along axis for important variable. (c) Axes represent two important variables, where neuroimaging similarity to S^+ is with respect to multivariate distances.

Table 1

Ten most significant ROIs under joint variance (JVAR) test

Greater significance and lower p-values under JVAR test implies that a ROI better characterizes S^+ (n=37) as a group relative to S^- (n=442) in thickness, area, and volume measurements simultaneously. (a) One-sided F-tests for ratio of variances, with 36 and 441 degrees of freedom, between the groups were performed for thickness, area, and volume variables of each respective ROI. (b) The p-values for the JVAR test statistic $T_j = 1 - \Pi_{k=1}^{3}(1-p_{jk})$ were generated through permutation distributions, as described in the Methods section. Significance of all 66 ROIs is provided in Supplementary Table 9. (c) FDR-adjusted p-values are over JVAR tests for 66 ROIs. (d) The left temporal pole had a FDR-adjusted p-value less than 0.05.

Rank	Region of Interest	Ratio of Variances F-test[a]: p_{jk}			Joint variance test[b] of $\mathcal{H}_{0j}$		
		Thickness	Area	Volume	Statistic T_j	p-value	FDR[c]
1	Left Temporal Pole [d]	0.044	0.014	0.002	0.0582	0.0005	0.035
2	Left Lingual	0.002	0.033	0.006	0.0403	0.0016	0.052
3	Right Middle Temporal	0.080	0.228	0.093	0.3558	0.0266	0.524
4	Left Caudal Anterior Cingulate	0.028	0.235	0.109	0.3381	0.0318	0.524
5	Right Rostral Anterior Cingulate	0.029	0.074	0.285	0.3576	0.0427	0.563
6	Left Transverse Temporal	0.031	0.053	0.385	0.4352	0.0533	0.586
7	Right Transverse Temporal	0.047	0.286	0.220	0.4686	0.0797	0.703
8	Left Entorhinal	0.065	0.063	0.372	0.4499	0.0852	0.703
9	Left Fusiform	0.286	0.256	0.244	0.5979	0.1211	0.802
10	Right Medial Orbitofrontal	0.016	0.460	0.284	0.6193	0.1409	0.802

Table 2

Models of learning performance for LP status and NS-LP score

Model A: $Y_i = a_0 + a_Z Z_i$; Model B: $Y_i = b_0 + b_L L_i$; Model C: $Y_i = c_0 + c_L L_i + c_Z Z_i$; Learning performance score Y_i is average of working memory, episodic memory, vocabulary, and reading exams. Each exam score was scaled to zero mean and unit variance after controlling for age, sex, household income, parents' education, and first two PCs of genotype. Indicator $L_i=1$ for positive LP diagnosis and $L_i=0$ otherwise. NS-LP score $Z_i := Z_i^v$ for $v = 0.01$. (a) F-test for respective models; degrees of freedom (df) are 1 and 477 for Models A and B, and df are 2 and 476 for Model C. (b) F-test for respective coefficients with 1 and 477 df's. (c) Model C's fit is 44% better than Model B as determined by percentage increase in adjusted R^2.

Summary of model fits

	Adjusted R^2	R^2	p-value [a]
Model A	0.024	0.026	3.7E-04
Model B [c]	0.036	0.038	1.6E-05
Model C [c]	0.052	0.056	9.7E-07

Model estimates

	beta	T	p-value [b]
Model A			
a_0	0.57	3.50	5.1E-04
a_Z	−0.36	−3.56	4.1E-04
Model B			
b_0	0.04	1.21	0.23
b_L	−0.49	−4.37	2.0E-05
Model C			
c_0	0.51	3.18	1.5E-03
c_L	−0.44	−3.85	1.3E-04
c_Z	−0.31	−3.01	2.8E-03

Table 3

Six SNPs on *AGBL1* gene (chromosome 15) significant in GWAS of learning performance. Tests of $\mathscr{J}_0^{(g)}$ and $\mathscr{M}_0^{(g)}$ were performed for G×R and Main effect models, respectively, for 488,200 SNPs of the learning performance scores over the PING sample (n=479). Both models included covariates for status of learning problem diagnosis and the first two PCs of genotype data. Learning performance scores were, by construction, uncorrelated with age, sex, household income, and parental education. Tests of $\mathscr{M}_{0;PNC}^{(g)}$ were performed for main effect models of WRAT scores over the PNC sample (n=2,327). Covariates there included age, sex, status of learning problem diagnosis, and genotyping platform (see Methods section). (a) F-tests of $\mathscr{J}_0^{(g)}$ were performed with statistic having 2 and 472 df. F-tests of $\mathscr{M}_0^{(g)}$ were performed with statistic having 1 and 474 df. F-tests of $\mathscr{M}_{0;PNC}^{(g)}$ were performed with statistic having 1 and 2,322 df. (b) FDR-adjustments were made to the p-values from 976,400 tests performed in the PING sample (n=479). (c) Information gain are for respective G×R and Main effect models over the PING sample (n=479). It represents the increase in adjusted R^2 for the respective model fit over that of the non-genetic base model including only the covariates. (e) Minor allele frequency is over the PING sample (n=479). (f) SNPs were genotyped with the Illumina Human660W-Quad Beadchip array and their positions correspond to HG build 19. See Supplementary Table 11 for minor alleles for the respective SNPs.

| SNP | MAF[e] (Position[f]) | p-value[a] (FDR[b]) | | | Information gain[c] Model | | Parameter estimates (standard error) G×R model | | | Main effect model | |
		$\mathscr{J}_0^{(g)}$	$\mathscr{M}_0^{(g)}$	$\mathscr{M}_{0;PNC}^{(g)}$	G×R	Main	α_g	β_g	γ_g	θ_g	$\theta_{g;PNC}$
rs11633708	0.47 (87050708)	3.5E-08 (0.020)	2.3E-06 (0.24)	7.0E-04	7.89	4.27	-0.81 (0.18)	-0.90 (0.22)	0.44 (0.14)	-0.20 (0.04)	-0.100 (0.030)
rs4243100	0.45 (87037559)	4.0E-08 (0.020)	6.5E-05 (0.90)	4.6E-04	7.84	3.01	0.24 (0.17)	1.03 (0.21)	-0.54 (0.13)	0.17 (0.04)	0.102 (0.029)
rs11853362	0.44 (87057124)	1.7E-07 (0.034)	7.7E-05 (0.90)	6.9E-04	7.29	2.95	-0.90 (0.18)	-1.00 (0.22)	0.51 (0.14)	-0.15 (0.04)	-0.100 (0.029)
rs11632104	0.45 (87049894)	2.0E-07 (0.034)	3.0E-04 (0.97)	3.0E-04	7.23	2.44	0.24 (0.17)	1.00 (0.22)	-0.53 (0.13)	0.15 (0.04)	0.105 (0.029)
rs4887473	0.45 (87053206)	2.0E-07 (0.034)	3.0E-04 (0.97)	3.0E-04	7.23	2.44	0.24 (0.17)	1.00 (0.22)	-0.53 (0.13)	0.15 (0.04)	0.106 (0.029)
rs2011905	0.45 (87053846)	2.1E-07 (0.034)	3.1E-04 (0.97)	3.3E-04	7.21	2.42	0.24 (0.17)	1.00 (0.22)	-0.53 (0.13)	0.15 (0.04)	0.105 (0.029)