Published in final edited form as:

Genet Epidemiol. 2018 April ; 42(3): 265–275. doi:10.1002/gepi.22111.

Whole Genome Association Study of Brain-wide Imaging Phenotypes: A Study of the PING Cohort

Canhong Wen^{1,2,3}, Chintan M. Mehta¹, Haizhu Tan^{1,4}, and Heping Zhang^{1,*}

¹Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA

²Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, China

³Department of Statistical Science, School of Mathematics, Sun Yat-Sen University, Guangzhou, China

⁴Department of Physics and Computer Applications, Shantou University Medical College, Shantou, China

Abstract

Neuropsychological disorders have a biological basis rooted in brain function, and neuroimaging data are expected to better illuminate the complex genetic basis of neuropsychological disorders. Because they are biological measures, neuroimaging data avoid biases arising from clinical diagnostic criteria that are subject to human understanding and interpretation. A challenge with analyzing neuroimaging data is their high dimensionality and complex spatial relationships. To tackle this challenge, we introduced a novel distance covariance tests that can assess the association between genetic markers and multivariate diffusion tensor imaging measurements, and analyzed a genome wide association study (GWAS) dataset collected by the Pediatric Imaging, Neurocognition, and Genetics (PING) study. We also considered existing approaches as comparisons. Our results showed that, after correcting for multiplicity, distance covariance tests of the multivariate phenotype yield significantly greater power at detecting genetic markers affecting brain structure than standard mass univariate GWAS of individual neuroimaging biomarkers. Our results underscore the usefulness of utilizing the distance covariance to incorporate neuroimaging data in GWAS.

Keywords

distance covariance; genome wide association studies; multivariate analysis; neuroimaging analysis

1. INTRODUCTION

With recent advances in neuroimaging and genotyping technology, the field of imaging genetics has emerged as a promising avenue for identifying genetic variants associated with

*Correspondence Author: Heping Zhang, Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520, USA; Telephone: (203) 785-5185; heping.zhang@yale.edu.

brain phenotypes of complex diseases, such as schizophrenia (Samartzis et al. (2014) and reference therein). Instead of directly mapping variants in the genome to clinical or behavioral traits, imaging genetics relates them to quantitative neuroimaging biomarkers. These biomarkers are believed to possess a more direct link to genetic variation than diagnostic measures based on cognitive or clinical assessments and thus potentially offer greater power in inferring associations between genetic variants and various neurological or psychiatric illness (Cannon and Keller, 2006).

The joint analysis of neuroimaging and genetic biomarkers is complicated by a high dimensionality of variables observed in each modality. Focusing on candidate imaging phenotypes and/or candidate genetic variants can help avoid statistical issues, namely the multiple testing problem, along with computational issues that arise due to the curse of dimensionality (Vounou et al., 2010; Liu and Calhoun, 2014). However, identifying candidate genetic variants and imaging phenotypes requires strong and robust scientific evidence. Another strategy is through a genome-wide association study (GWAS) of brain-wide imaging phenotypes. Stein et al. (2010) introduced vGWAS, a univariate analysis to the pair of each voxel and each single nucleotide polymorphism (SNP), and adjusted for multiple comparisons to control the false positive rate. With high dimensionality in both genotypes and imaging phenotypes, Stein et al. (2010) failed to find any significant result in their analysis after correcting for multiple comparisons. Furthermore, vGWAS is computationally intensive because it typically runs genome-wide associations with 100,000 voxels or more in the whole brain. To reduce the number of tests and enhance the power of detecting genetic variants, one could take advantage of the spatial information in images and/or combining the effect of multiple genetic variants. Shen et al. (2010) proposed GWAS with regions of interest (ROIs) rather than voxels in imaging phenotypes. Hibar et al. (2011) reduced the dimensionality of the genome by considering voxel wise association with each gene. Ge et al. (2012) proposed using cluster-wise inference to detect spatially distributed signals based on random field theory. Huang et al. (2015) proposed a fast voxel-wise GWAS method with a pre-screening step that considered all the voxels as a whole.

The Pediatric Imaging, Neurocognition, and Genetics (PING) was a study undertaken to understand the genetic basis of individual differences in brain morphology and cognitive attributes. PING assessed measures of brain connectivity using diffusion tensor imaging (DTI), which have recently become a popular way to investigate white matter micro structural changes that are sensitive to brain development and degeneration. These measures included a total of 336 imaging phenotypes, with eight different metrics for 42 fiber tracts across the brain. We performed a mass-univariate GWAS analysis for each of these 336 imaging phenotypes and used hierarchical clustering to identify association patterns between top SNPs and top imaging phenotypes at various levels of significance. However, no association test between single imaging phenotype and single SNP survived after correcting for multiple testing. Thus we proposed and performed a multivariate approach for GWAS using distance covariance tests between each genotyped SNP and all 42 T2-weighted imaging measurements over the whole brain. Unlike the mass-univariate analysis, the distance covariance test captures the spatial correlation in multivariate imaging phenotypes. We also examined associations between SNPs having highest significance in the GWAS and cognitive scores with models that either included and excluded the multivariate imaging

phenotypes as covariates. Adjusting for the imaging phenotypes enhanced the association signals between these SNPs and cognitive scores.

2. MATERIALS AND METHODS

The analysis presented in this paper is on data obtained from the PING Study web portal (<http://ping.chd.ucsd.edu>). PING was launched in 2009 by the National Institute on Drug Abuse (NIDA) and the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) as a two-year project of the American Recovery and Reinvestment Act. The primary goal of PING was to create a data resource of highly standardized and carefully curated magnetic resonance imaging (MRI) data, comprehensive genotyping data, and developmental and neuropsychological assessments for a large cohort of developing children aged 3 to 20 years. The scientific aim of the project is, by openly sharing these data, to amplify the power and productivity of investigations of healthy and disordered development in children, and to increase understanding of the origins of variation in neurobehavioral phenotypes. For up-to-date information, see <http://ping.chd.ucsd.edu/>. In this study, we focus on the 671 individuals in PING with complete genetic data, brain DTI measurements, cognitive assessments, and a few specific covariates including the genetic ancestry factor (GAF), scanner, sex, age, and socioeconomic status (household income and the highest parental education).

2.1 Diffusion tensor imaging (DTI)

DTI is a neuroimaging modality for examining structural connectivity in the human cortex by tracking movement (diffusivity) in water molecules, which traverse with the greatest speed along white matter fiber tracts. More precisely, diffusivity is the three-dimensional gradient of water molecules measured by attenuation of the T2-weighted MRI signal and is represented by a 3×3 positive definite matrix, or tensor. Considered in our analysis are four common metrics quantifying diffusivity at each voxel, which are derived from the eigenvalues of the tensor matrix. Le Bihan et al. (2001) included mean, longitudinal and transverse diffusivity (MD, LD and TD), along with fractional anisotropy (FA).

The FA, MD, LD, and TD measures at the voxel level were aggregated into 35 regions of interest (ROIs) that correspond to known anatomical white matter tracts. These tracts were identified and labeled with an automated atlas-based method based on a probabilistic atlas of fiber tract locations and orientations (Hagler et al., 2009). Four other MRI-derived summary measures for these 35 ROIs considered in our analysis include normalized and unnormalized T2-weighted MRI intensities (T2w and b0N), normalized T1-weighted intensity, and fiber tract volume. Each respective measure for the white matter fiber tract in 35 ROIs was aggregated for the entire cortex, left hemisphere, and right hemisphere. Four additional aggregate measures were considered for the left or right hemispheres when the corpus callosum is excluded and the left or right fornix when the fimbria is excluded.

With eight measures of white fiber tracts in the 35 ROIs and seven aggregated ROIs, we considered a total of 336 DTI phenotypes. PING DTI imaging techniques, data acquisition, and analyses are discussed briefly below. In depth descriptions can be found elsewhere (Fjell et al., 2012; Eicher et al., 2013, 2015).

2.2 Genotypes

PING participants were genotyped with the Illumina Human660W-Quad platform. We restricted our analysis to SNPs each meeting each of the following quality control criteria: genotype call rate < 90%, significant deviation from the Hardy-Weinberg equilibrium where p -value < 0.0001, minor allele frequency < 0.05, or available from < 90% of the participants. There were 487,588 autosomal SNPs meeting these quality control criteria. The final count of SNPs considered in our analysis on each chromosome is reported in Table 1.

2.3 Mass-univariate GWAS

As described in Section 2.1, we have 336 selected imaging phenotypes. We first conducted a separate GWAS for each of the 336 imaging phenotypes. We used a linear regression model to test the additive effect of each of the 487,588 SNPs on a given imaging phenotype, while adjusting for genetic ancestry factor (GAF), scanner, sex, age, household income, and parental education based on previous studies (Mehta et al., 2017). The empirical p -values were based on the Wald statistic. A threshold of $P < 3.05 \times 10^{-10}$ was considered as significant after the Bonferroni correction, controlling for a total of 487,588 SNPs and 336 imaging phenotypes. Hierarchical clustering and heat maps were used to visualize the GWAS results.

To better understand the mass-univariate GWAS results and set the stage for further analysis, we employed the heat map and hierarchical clustering approaches (Yau, 2014) to visualize associations between SNPs and the selected imaging phenotypes. The colors in the heat maps reflect $\log_{10}(p)$ of the corresponding associations. We also included dendrograms in the heat maps to represent hierarchical clustering for associations between SNPs and imaging phenotypes. On each heat map, associations with p -values less than 1.0×10^{-7} between SNPs and imaging phenotypes were marked with an “x” to facilitate visual evaluation of the results. The color bar on the left hand side of the heat maps encodes the chromosome IDs for the corresponding SNPs.

2.4 Multivariate phenotype GWAS

We used the distance covariance statistic (Szekely et al., 2007) to test associations between each SNP and the multivariate phenotype consisting of 42 selected T2w measures, the initial 336 imaging phenotypes. Without this screening step, we would include too much noise in the multivariate phenotype and lower the power (Ramdas, 2015). We should note that the T2w image is known to be superior to other measures in detecting abnormality in cerebrospinal fluid, which is related to a variety of neurological disorders and neuro-degeneration (Douaud et al., 2013; Amlie et al., 2014). Prior to the distance covariance testing, each of the T2w measures was adjusted for GAF, scanner, sex, and age. A Manhattan plot and a quantile-quantile (Q-Q) plot were used to visualize the GWAS results.

The distance covariance statistic is, most generally, evaluated between two random vectors X and Y of dimensions p and q . In the present study, for the i -th participant, let $Y_i = (Y_{i,1}, \dots, Y_{i,42})$ be the vector of 42 T2w measurements and X_i the number of minor alleles in SNP g . Let $(X, Y) = \{(X_i, Y_i): i = 1, \dots, n\}$ be the observed data. Here, $n = 671$. As Székely et al.

(2007) and Székely et al. (2009) discussed, the sample version of distance covariance between X and Y takes the form

$$v_n = \widehat{dCov}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{k,l=1}^n A_{kl} B_{kl},$$

where, for $k, l = 1, \dots, n$,

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot},$$

and

$$a_{kl} = \|X_k - X_l\|, \quad \bar{a}_{k\cdot} = \frac{1}{n} \sum_{l=1}^n a_{kl}, \quad \bar{a}_{\cdot l} = \frac{1}{n} \sum_{k=1}^n a_{kl}, \quad \bar{a}_{\cdot\cdot} = \frac{1}{n^2} \sum_{k,l=1}^n a_{kl},$$

where $\|X_k - X_l\| = \sqrt{(X_{k,1} - X_{l,1})^2 + \dots + (X_{k,42} - X_{l,42})^2}$.

Similarly, define $b_{kl} = \|Y_k - Y_l\|$ and $B_{kl} = b_{kl} - b_{k\cdot} - b_{\cdot l} + \bar{b}_{\cdot\cdot}$. A version of the adaptive permutation was used to obtain the empirical p-value (Besag and Clifford, 1991). The basic idea is to run more permutations for smaller p-values as we are most interested in small p-values. Since the multivariate phenotype is considered as a single response, the threshold of significance for p-values was set to be $P = 0.05/487,588 = 1.03 \times 10^{-7}$.

2.5 Alternate multivariate models

For comparison, we considered three alternative approaches that can accommodate the 42 T2-weighted DTI intensity measures as the multivariate response.

First, let us consider the following model:

$$Y_i = \mu_g + b_g X_{ig} + \varepsilon_{g;i},$$

where μ_g and b_g are 42-dimensional vectors for location and genetic effect parameters, and $\varepsilon_{g;i} \sim \mathcal{N}_{42}(0, \Sigma_g)$ are iid noise vectors of dimension 42 that have a multivariate normal distribution with zero mean and unspecified covariance. We performed a Pillai's test of the null hypothesis $b_g = 0$.

Next, we performed principal component (PC) analysis of the Y_i 's. Let $Y_i^J \triangleq \hat{V}^J Y_i$ denote the vector of scores on the first J PCs, where $\hat{V}^J$ is the $(J \times 42)$ -dimensional matrix whose rows contain estimated loadings on the respective J PCs. After observing that the first six components explained 90% of variation, we performed tests of genetic effect in a

multivariate model of the first six PC scores, evaluated as $Y_i^J = \mu_g^J + b_g^J X_{ig}^J$, through a Pillai's test of $b_g^J = 0$ for $J = 6$.

Finally, we considered each of the first 6 PC scores as a univariate phenotype and performed a univariate GWAS.

Again, the threshold of significance for p-values was set at $P = 1.03 \times 10^{-7}$, as above.

2.6 Link with cognitive measures

Participants in the PING study received cognitive assessments with a battery of eight exams in the NIH Toolbox for Cognition Battery (NTCB). These exams measure an individual's performance across eight cognitive domains; see Table 2. The content and administration of the exams are described at greater length in Akshoomoff et al. (2014).

To characterize the joint link between genetics, neuroimaging, and clinical outcomes, we examined effects from the five SNPs having greatest significance in the refined GWAS on cognitive traits. We tested the effect from these SNPs in two linear models of scores on eight cognitive exams: one included the 42 T2w measurements as covariates while the other excluded them. As this is a candidate gene approach, we used a nominal significance level of $P = 0.05/8/5 = 0.00125$ as a rough threshold. Furthermore, we also constructed two composite cognitive scores through principal component analysis and examined their associations with the five SNPs. In this case, the significance level was set at $P = 0.05/2/5 = 0.005$.

3. RESULTS

3.1 Mass-univariate GWAS

In the mass-univariate GWAS, an association test was performed between each of 487,588 SNPs and 336 imaging phenotypes, giving rise to approximately 1.6×10^{-8} SNP-imaging biomarker pairs. Among them, associations for 48 pairs representing 24 unique SNPs were significant at the 10^{-7} level (see Table 3). No SNP surpassed the GWAS threshold of 3.05×10^{-10} .

The heat map in Figure 1 illustrates the 48 associations meeting significance threshold of 10^{-7} (see blocks labeled with "x"). The horizontal axis corresponds to imaging phenotypes, with those of the same type (i.e., T2-weighted) clustered together. The vertical axis corresponds to SNPs. Six SNPs from the PEMT gene had significant associations with the fractional anisotropy for the right fornix excluding the fimbria. Their grouping together in Figure 1, suggests an increased likelihood of linkage disequilibrium (LD) effects.

3.2 Multivariate phenotype GWAS

We next performed a multivariate phenotype GWAS for the 42 T2w measures, as they predominated the heat map in Figure 1, using distance covariance tests. The twenty SNPs with highest significance in this GWAS are reported in Table 4. Manhattan and Q-Q plots from this GWAS are shown in Figure 2. The Q-Q plots reveal no evidence of systematic bias due to factors such as a strong population stratification, cryptic relatedness, or genotyping

artifacts. The p-values in the upper tail of the distribution do show, however, a clear deviation from the null distribution of no association, suggesting the presence of significant associations between those SNPs and imaging phenotypes.

Five SNPs were identified as having an association with the multivariate imaging phenotypes at the 10^{-5} significance level. In particular, we found that rs6910079 was significantly associated with the 42 T2w measures ($P = 5.2 \times 10^{-8}$). It is located in an intergenic region between the QKI and LOC728316 genes on chromosome 6. Note that the minimum p-value with individual imaging phenotypes in the mass-univariate analysis for SNP rs6910079 is 1.4×10^{-8} , which is much larger than $5 \times 10^{-8}/42 = 1.19 \times 10^{-9}$ – the Bonferroni-corrected significance threshold if we only consider the 42 T2w measures in the mass-univariate analysis. SNPs, rs1565654 and rs12229466, in the same gene TMEM132D on chromosome 12, may be potentially associated with the T2w measures in the DTI ROIs (rs12229466: $P = 2.1 \times 10^{-6}$; rs1565654: $P = 2.4 \times 10^{-6}$). TMEM132D is a transmembrane protein coding gene expressed in neurons and colocalized with actin filaments that putatively function as a cell-surface marker for differentiation of oligodendrocyte differentiation (Nomoto et al., 2003). This gene encodes a single-pass type I membrane protein belonging to the TMEM132 protein family that is highly expressed in the cortical regions of the human and mouse brain.

To gain further insight into the identified SNPs by distance covariance based GWAS, we presented a heat map of SNP association results from mass-univariate GWAS for the SNPs in Table 4 and the 42 T2w measurements in Figure 3. The pattern in this heat map follows from correlations between the imaging variables.

We then performed multivariate analysis of variance of the 42 imaging phenotypes in linear models for the 5 SNPs in Table 4. Although MANOVA indicated stronger significant associations when projecting the imaging variables onto their first six principal components, none of the associations survived after the multiple comparison correction. This indicated the limitation of MANOVA in application with a moderate number of phenotypes. Associations remained strong when we project the phenotypes onto the first PC only, which partially why the distance covariance is advantageous. However, the p-values of a Pillai's test between the first PC and each SNP was markedly larger than in the distance covariance test. Thus, the distance covariance test is superior to the principal component analysis in the current study.

3.3 Link with cognitive measures

Recall that so far we have not used cognitive exams in the GWAS. As the last step, we examined the associations between the five SNPs identified in Section 3.2 and the eight cognitive exams.

Principal component analysis of the eight cognitive exams indicated that the first two factors explain 79.51% of variation in the scores. Exam loadings for the first factor, which explained 70.92% of variation, were all in the same direction suggesting that it represents an overall performance score. On the second factor, which added 8.58% ($= 79.51\% - 70.92\%$) to the

explained variance beyond the first factor, the loadings were dominated by the attention and flanker exams.

In Table 5, we presented the p-values of the associations between the five most significant SNPs in Section 3.2 and the eight individual cognitive scores individually or collectively through their first two principal components. While assessing those associations, we excluded or included the 42 T2w measurements as the covariates for comparison.

When the T2w measurements were included in the model (the bottom one of the two p-values), several association results became much more significant. For example, the p-values for the associations between pspac score and rs1565654 decreased to 0.08 from 0.31 when the 42 T2w were adjusted for, and likewise for rs12229466.

The first factor indicated a significant association with rs7443752 ($P = 0.005$). Even though the p-value between LS and rs7443752 was 0.003, it did not factor in the fact that LS is only one of the eight individual scores. Moreover, the association between the second factor and two SNPs (rs12229466 and rs946263) became potentially significant ($p = 0.03$ and 0.04) after adjusting for the 42 T2w measurements.

4. DISCUSSION

Clinical assessments of many neuropsychological disorders and traits have complex genetic architectures. Because variations in brain structure and function are believed to be closer to underlying biological etiologies of such disorders than genetic variation, quantitative biomarkers derived from neuroimaging are thus suitable endophenotypes. They offer tremendous potential for early diagnoses, monitoring of treatment, and understanding pathophysiology from genetic variants in many neurological and psychiatric disorders including schizophrenia, autism, and Alzheimer's disease.

We conducted genome-wide association analysis of measures for white fiber tracts in the brain with two different approaches. We first used a mass-univariate analysis where each of 336 neuroimaging biomarkers were examined separately. This approach failed to yield any significant association at the imaging phenotype and genome-wide levels ($P < 0.05/(487, 588 \times 336) = 3.05 \times 10^{-10}$). The challenge is the excessive number of association tests. However, the top-ranked findings suggested that multivariate imaging phenotypes had potentially significant associations with some SNPs, which suggested that the degrees of freedom in neuroimaging data were, in effect, much smaller than their high dimensionality implied. Indeed, spatial contiguity properties of neuroimaging data imply that they are better viewed as continuous functions in space (Medland et al., 2014).

Based on our results in the mass-univariate analysis, we treated all the imaging data as a single phenotype and utilized a test of multivariate independence, the distance covariance test, to identify significant associations. The lowest p-value in a GWAS of the multivariate phenotype ($P = 5.2 \times 10^{-8}$) was for SNP rs6910079, which lies in the intergenic regions between the QKI and LOC728316 genes. QKI is a brain-specific gene linked to myelination and oligodendrocyte differentiation that may play a role in schizophrenia (Åberg et al., 2006). We also found nominal evidence of association with several SNPs in

gene TMEM132D on chromosome 12q24.33. Although little is known about the molecular function of the TMEM132D, this gene has been implicated in abnormal emotional behavior and management of mental disorder, especially for fear and anxiety-related behavior (Erhardt et al., 2011; Mick et al., 2011; Quast et al., 2012; Haaker et al., 2014). For example, one family-based genome wide study found rs11060369 in TMEM132D gene as being highly associated ($P = 2.3 \times 10^{-5}$) with the Child Behavior Checklist profile comprising anxious and depression subscales Mick et al. (2011). Erhardt et al. (2011) also identified SNPs from TMEM132D as being significant in a genome-wide association study for panic disorder.

To further explore the biological function of the top-ranked SNPs, multiple regression analysis was performed to associate these SNPs to cognitive traits assessed from a battery of exams. We considered models of cognitive scores that respectively included and excluded the multivariate imaging phenotypes. Adjusting for the imaging phenotypes generally enhanced the association signals between the SNPs and cognitive scores, whether the scores were considered individually or through composites defined through PCA. This observation underscores the potential usefulness of including imaging phenotypes in the genetic association studies of cognitive scores.

One limitation of our analysis is that the sample size of our brain imaging data is not large enough. Thus it is not surprising that the mass-univariate and some multivariate analyses did not produce significant results. This is the case for similar studies (e.g., Bi et al. 2017). With the Brain Initiative and other ongoing efforts we expect more and larger imaging-genetic datasets to be available in the near future. In this regard, we believe our proposed approach is ahead of the curve and will become more useful in the field.

Another limitation lies in the re-use of the same data in the univariate analysis and the multivariate analysis on a subset of the imaging phenotypes. It would be ideal to have two datasets: one for screening and one for further testing. This may become possible in the future, but at the present, the effort of collecting and processing imaging data has limited the availability of large and similarly designed imaging-genetic datasets. Actually, the univariate approach can be viewed as a screening step, a common practice in statistics. Furthermore, we would include too much noise if considering all the imaging phenotypes in the multivariate approach. Thus we selected a small subset of the imaging phenotypes to perform the multivariate analysis.

Acknowledgments

This research is supported in part by grants R01 DA016750 from the National Institute on Drug Abuse (HZ), T32 MH 14235 from the National Institute of Mental Health (CM). Data collection and sharing for this project was funded by the PING (National Institutes of Health Grant RC2DA029475, RC2MH089983, and RC2MH089924). PING is funded by the National Institute on Drug Abuse and the Eunice Kennedy Shriver National Institute of Child Health and Human Development. PING data are disseminated by the PING Coordinating Center at the Center for Human Development, University of California, San Diego. The key investigators of PING include Hauke Bartsch, Anders M. Dale, Anthony Gamst, Connor McCabe, Erik Newman, Joshua Kuperman, Mark Appelbaum, Natacha Akshoomoff, Terry L. Jernigan, Wesley Thompson, Peter Van Zijl, Stewart Mostofsky, Walter Kaufmann, Bruce Rosen, Tal Kenet, Alisa Powers, B.J. Casey, Erika J. Ruberry, Nicholas J. Schork, Cinnamon Bloss, Sarah Murray, David Amaral, Elizabeth Sowell, Brian Keating, Linda Chang, Thomas Ernst, David Kennedy, Jean Frazier, and Jeffrey R. Gruen.

References

- Åberg K, Saetre P, Jareborg N, Jazin E. Human qki, a potential regulator of mrna expression of human oligodendrocyte-related genes involved in schizophrenia. *Proceedings of the National Academy of Sciences*. 2006; 103(19):7482–7487.
- Akshoomoff N, Newman E, Thompson WK, McCabe C, Bloss CS, Chang L, Amaral DG, Casey B, Ernst TM, Frazier JA, et al. The nih toolbox cognition battery: Results from a large normative developmental sample (ping). *Neuropsychology*. 2014; 28(1):1. [PubMed: 24219608]
- Amlien IK, Fjell AM. Diffusion tensor imaging of white matter degeneration in Alzheimer's disease and mild cognitive impairment. *Neuroscience*. 2014; 276:206–215. [PubMed: 24583036]
- Besag J, Clifford P. Sequential Monte Carlo p-values. *Biometrika*. 1991:301–304.
- Cannon TD, Keller MC. Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol*. 2006; 2:267–290. [PubMed: 17716071]
- Douaud G, Menke RA, Gass A, Monsch AU, Rao A, Whitcher B, ... Smith S. Brain microstructure reveals early abnormalities more than two years prior to clinical progression from mild cognitive impairment to Alzheimer's disease. *Journal of Neuroscience*. 2013; 33(5):2147–2155. [PubMed: 23365250]
- Eicher JD, Montgomery AM, Akshoomoff N, Amaral DG, Bloss CS, Libiger O, Schork NJ, Darst BF, Casey B, Chang L, et al. Dyslexia and language impairment associated genetic markers influence cortical thickness and white matter in typically developing children. *Brain imaging and behavior*. 2015:1–11. [PubMed: 25724689]
- Eicher JD, Powers NR, Miller LL, Akshoomoff N, Amaral DG, Bloss CS, Libiger O, Schork NJ, Darst BF, Casey B, et al. Genome-wide association study of shared components of reading disability and language impairment. *Genes, Brain and Behavior*. 2013; 12(8):792–801.
- Erhardt A, Czibere L, Roeske D, Lucae S, Unschuld PG, Ripke S, Specht M, Kohli MA, Kloiber S, Ising M, et al. Tmem132d, a new candidate for anxiety phenotypes: evidence from human and mouse studies. *Molecular psychiatry*. 2011; 16(6):647–663. [PubMed: 20368705]
- Fjell AM, Walhovd KB, Brown TT, Kuperman JM, Chung Y, Hagler DJ, Venkatraman V, Roddey JC, Erhart M, McCabe C, et al. Multimodal imaging of the self-regulating developing brain. *Proceedings of the National Academy of Sciences*. 2012; 109(48):19620–19625.
- Ge T, Feng J, Hibar DP, Thompson PM, Nichols TE. Increasing power for voxel-wise genome-wide association studies: the random field theory, least square kernel machines and fast permutation procedures. *Neuroimage*. 2012; 63(2):858–873. [PubMed: 22800732]
- Haaker J, Lonsdorf T, Raczka K, Mechias M, Gartmann N, Kalisch R. Higher anxiety and larger amygdala volumes in carriers of a tmem132d risk variant for panic disorder. *Translational psychiatry*. 2014; 4(2):e357. [PubMed: 24495968]
- Hagler DJ, Ahmadi ME, Kuperman J, Holland D, McDonald CR, Halgren E, Dale AM. Automated white-matter tractography using a probabilistic diffusion tensor atlas: Application to temporal lobe epilepsy. *Human brain mapping*. 2009; 30(5):1535–1547. [PubMed: 18671230]
- Hibar DP, Stein JL, Kohannim O, Jahanshad N, Saykin AJ, Shen L, Kim S, Pankratz N, Foroud T, Huentelman MJ, et al. Voxelwise gene-wide association study (vgenewas): multivariate gene-based association testing in 731 elderly subjects. *Neuroimage*. 2011; 56(4):1875–1891. [PubMed: 21497199]
- Huang M, Nichols T, Huang C, Yang Y, Lu Z, Feng Q, Knickmeyer RC, Zhu H, Initiative ADN, et al. FVGWAS: Fast voxelwise genome wide association analysis of large-scale imaging genetic data. *NeuroImage*. 2015
- Le Bihan D, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N, Chabriat H. Diffusion tensor imaging: concepts and applications. *Journal of magnetic resonance imaging*. 2001; 13(4):534–546. [PubMed: 11276097]
- Liu J, Calhoun VD. A review of multivariate analyses in imaging genetics. *Frontiers in neuroinformatics*. 2014:8. [PubMed: 24600385]
- Medland SE, Jahanshad N, Neale BM, Thompson PM. Whole- genome analyses of whole-brain data: working within an expanded search space. *Nature neuroscience*. 2014; 17(6):791–800. [PubMed: 24866045]

- Mehta CM, Gruen JR, Zhang H. A method for integrating neuroimaging into genetic models of learning performance. *Genetic Epidemiology*. 2017; 41(1):4–17. [PubMed: 27859682]
- Mick E, McGough JJ, Loo S, Doyle AE, Wozniak J, Wilens TE, Smalley SL, McCracken JT, Biederman J, Faraone SV. Genome-wide association study of the child behavior checklist dysregulation profile. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2011; 50(8):807. [PubMed: 21784300]
- Nomoto H, Yonezawa T, Itoh K, Ono K, Yamamoto K, Oohashi T, Shiraga F, Ohtsuki H, Ninomiya Y. Molecular cloning of a novel transmembrane protein molt expressed by mature oligodendrocytes. *Journal of Biochemistry*. 2003; 134(2):231–238. [PubMed: 12966072]
- Quast C, Altmann A, Weber P, Arloth J, Bader DM, Heck A, Pfister H, Mullermyhsok B, Binder Elisabeth B, AE. Rare variants in tmem132d in a case control sample for panic disorder. *American Journal of Medical Genetics*. 2012; 159B(8):896–907. [PubMed: 22911938]
- Ramdas, A., Reddi, SJ., Póczos, B., Singh, A., Wasserman, LA. On the Decreasing Power of Kernel and Distance Based Nonparametric Hypothesis Tests in High Dimensions. AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence; 2015. p. 3571-3577.
- Samartzis L, Dima D, Fusar-Poli P, Kyriakopoulos M. White matter alterations in early stages of schizophrenia: a systematic review of diffusion tensor imaging studies. *Journal of Neuroimaging*. 2014; 24(2):101–110. [PubMed: 23317110]
- Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, Foroud T, Pankratz N, Moore JH, Sloan CD, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in mci and ad: A study of the adni cohort. *Neuroimage*. 2010; 53(3):1051–1063. [PubMed: 20100581]
- Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, et al. Voxelwise genome-wide association study (vGWAS). *Neuroimage*. 2010; 53(3):1160–1174. [PubMed: 20171287]
- Székely GJ, Rizzo ML, Bakirov NK, et al. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*. 2007; 35(6):2769–2794.
- Székely GJ, Rizzo ML, et al. Brownian distance covariance. *The annals of applied statistics*. 2009; 3(4):1236–1265.
- Vounou M, Nichols TE, Montana G, Initiative ADN, et al. Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. *Neuroimage*. 2010; 53(3):1147–1159. [PubMed: 20624472]
- Yau, C. R Tutorial with Bayesian Statistics Using OpenBUGS. 2014. r-tutor.com

Figure 1. Heat map of SNP associations with imaging phenotypes at the significance level of $p < 10^{-7}$. $-\log_{10}(p\text{-values})$ from the GWAS are color-mapped and displayed in the heat map. Heat map blocks labeled from “x” reach the significance level of $p < 10^{-7}$. Only top SNPs (associated with at least one imaging phenotype) and top imaging phenotypes (associated with at least one SNP) are included in the heat maps, and so each row (SNP) and column (imaging phenotype) has at least one “x” block. Dendrograms derived from hierarchical clustering are plotted for both SNPs and imaging phenotypes. The color bar in the bottom of the heat map codes the chromosome IDs for the corresponding SNPs.

Figure 2.
Manhattan and QQ-plots for GWAS results based on distance covariance test.

Figure 3. Heat map of SNP associations results from mass-univariate GWAS for the SNPs in Table 4 and the 42 T2w measurements. $-\log_{10}(\text{p-values})$ from each GWAS are color-mapped and displayed in the heat map. The heat map blocks labeled with “x” reach the significance level of $p < 10^{-7}$. The color bar on the bottom of the heat map codes the chromosome IDs for the corresponding SNPs.

Table 1

Final number of SNPs on each chromosome analyzed

CHR	# of SNPs	CHR	# of SNPs	CHR	# of SNPs	CHR	# of SNPs
1	37927	7	27228	13	18798	19	8772
2	40131	8	27686	14	16504	20	12638
3	33801	9	23899	15	14834	21	7499
4	29569	10	25985	16	15200	22	7433
5	30543	11	24244	17	13017		
6	32675	12	24223	18	14982	Total	487588

Table 2Description of exams in NIH Toolbox ^{*}

Test	Full Name	Ability	Subdomain
IBAM	Imitation-based Assessment of Memory Test	Episodic Memory	Episodic Memory
LS	List Sorting Working Memory Test	Working Memory	Working Memory
Vocab	Picture Vocabulary Test	Vocabulary Knowledge	Language
Reading	Oral Reading Recognition Test	Oral Reading Skill	Language
PSPAC	Pattern Comparison Processing Speed Test	Processing Speed	Processing Speed
Attention	Attention Test	Visual Attention	Attention
Flanker	Flanker Inhibitory Control Test	Inhibitory Control	Executive Function
DCCS	Dimensional Change Card Sort Test	Cognitive Flexibility	Executive Function

^{*} Taken from Table 3 in Askshoomoff, et al (2014)

Table 3

Information for the identified SNPs and their respective genes

SNP	CHR	Position	Gene	Left/Right Gene	# of DTI	P _{min}
rs9380890	6	39721216		KIF6/DAAAM2	8	1.84E-09
rs6910079	6	164369451		QKI/LOC728316	6	1.40E-08
rs17371566	1	180119349		CEP350/QSOX1	5	7.48E-09
rs2218495	2	193742078		PCGEM1/LOC645314	3	1.72E-09
rs186640	2	137323577		LOC648390/THSD7B	3	2.04E-08
rs12229466	12	130100796	TMEM132D		3	3.06E-08
rs7910211	10	81228378		ZCCHC24/TPRX1P1	2	1.50E-08
rs1565654	12	130104797	TMEM132D		2	1.70E-08
rs1108367	17	76558773	DNAH17		2	2.45E-08
rs2869617	7	68350360		STAG3L4/AUTS2	2	4.41E-08
rs1869966	4	155979144		LOC729902/NPY2R	1	1.09E-08
rs946263	1	203165381		CHI3L1/CHIT1	1	1.44E-08
rs2670049	8	57400730		PENK/RPL37P6	1	2.23E-08
rs8012197	14	59761192	DAAAM1		1	3.01E-08
rs2555625	5	174330193		LOC105377742/FLJ16171	1	5.31E-08
rs10191541	2	12901718		TRIB2/FAM84A	1	6.11E-08
rs4646372	17	17460180	PENT		1	6.72E-08
rs42444598	17	17472613	PENT		1	7.11E-08
rs4646350	17	17474154	PENT		1	7.11E-08
rs4646346	17	17478917	PENT		1	8.05E-08
rs137653	22	39732171		RPL3/SYNGR1	1	8.49E-08
rs4646365	17	17467783	PENT		1	9.28E-08
rs4646359	17	17470609	PENT		1	9.28E-08

Note: Information for the identified SNPs and their respective genes (i.e., the closest gene) at the significance level of $p < 10^{-7}$ is shown in this table, including SNP, chromosome (CHR), position, gene and left/right genes. In addition, the number of imaging phenotypes that are associated with each SNP at the significance level of $p < 10^{-7}$ and the minimum p-values from the 336 univariate GWAS are also shown. The SNPs are ordered according to the last two columns.

Table 4

The top 20 associated SNPs ranked by the permuted p-values from distance covariance test

SNP	CHR	Position	Gene (IG: Left/Right gene)	Distance Covariance	Test		
					MANOVA	F-test	on PC1
rs6910079	6	164369451	IG: QKI/LOC728316	5.20E-08	6.90E-02	9.00E-06	1.30E-07
rs12229466	12	130100796	TMEM132D	2.10E-06	1.10E-03	6.50E-06	1.70E-05
rs1565654	12	130104797	TMEM132D	2.40E-06	3.70E-03	4.20E-06	1.80E-05
rs7443752	5	70716555	IG: PMCHL2/BDP1	6.50E-06	7.30E-03	9.40E-05	2.10E-05
rs946263	1	203165381	IG: CH13L1/CHIT1	7.20E-06	2.30E-03	5.50E-05	7.40E-07
rs4237723	11	12638232	IG: PARVA/TEAD1	1.40E-05	1.00E-02	4.40E-04	7.10E-06
rs1435972	5	20124530	IG: CDH18/NUP50P3	1.60E-05	1.00E-01	2.60E-03	1.20E-04
rs7399129	12	130097213	TMEM132D	2.20E-05	4.20E-02	5.80E-04	5.70E-05
rs12607689	18	20016299	IG: CTAGE1/LOC646583	2.30E-05	9.90E-02	1.30E-03	1.00E-04
rs26055	5	111040812	IG: STARD4/C5orf13	2.60E-05	7.30E-02	1.40E-03	8.10E-06
rs4791284	17	65516649	PITPNCl	2.90E-05	7.00E-02	3.80E-03	4.40E-05
rs133796	22	44842135	LOC100131299	3.10E-05	3.30E-01	2.90E-03	5.00E-05
rs133795	22	44842354	LOC100131299	3.10E-05	2.40E-01	1.30E-03	3.30E-05
rs11064255	12	6669257	NOL1	3.20E-05	1.10E-01	3.90E-03	6.90E-05
rs766304	17	66590851	FAM20A	3.30E-05	1.10E-03	7.10E-04	6.40E-05
rs2099308	3	107585300	IG: BBX/LOC285205	3.60E-05	3.20E-02	1.80E-05	1.90E-05
rs210798	6	135514558	MYB	3.60E-05	1.30E-02	9.20E-06	5.80E-05
rs12676599	8	142745374	IG: MROH5/MIR1302-7	4.00E-05	1.20E-01	1.10E-05	3.20E-05
rs1561283	2	85965769	IG: GNLY/ATOH8	4.40E-05	2.10E-02	7.10E-03	4.50E-05
rs11077678	17	66626373	IG: FAM20A/ABCA8	4.40E-05	1.30E-03	3.50E-04	6.40E-05

Table 4 showed that the forty-two T2-weighted measures were used as a multivariate phenotype to identify the association with each single SNP. Associations were tested for these SNPs and the phenotypes in three models. MANOVA was performed on both the 42 T2-weighted measures and their projections on to the first principal components. Significance is under the Pillai's test. An F-test was performed for the association between each SNP and the first PC of the multivariate phenotype. The bold font indicates test having greatest significance for respective SNP.

Table 5

Association results between cognitive scores and the top 5 SNPs

SNP	IBAM	LS	PSPAC	Vocab	Reading	Flanker	Attention	DCCS	1st PC	2nd PC
rs6910079	0.43	0.48	0.56	0.49	0.14	0.23	0.53	0.27	0.25	0.91
rs12229466	0.62	0.42	0.96	0.91	0.4	0.09	0.18	0.15	0.16	0.34
	0.29	0.59	0.31	0.21	0.82	0.57	0.47	0.71	0.88	0.12
rs1565654	0.25	0.45	0.08	0.08	0.33	0.5	0.39	0.78	0.81	0.03
	0.59	0.71	0.31	0.43	0.47	0.36	0.25	0.91	0.57	0.23
rs7443752	0.6	0.84	0.07	0.19	0.94	0.28	0.19	0.89	0.57	0.07
	0.03	0.02	0.79	0.93	0.81	0.06	0.05	0.09	0.04	0.51
rs946263	0.008	0.003	0.61	0.36	0.31	0.02	0.02	0.05	0.005	0.82
	0.27	0.34	0.65	0.84	0.38	0.38	0.43	0.5	0.83	0.26
	0.23	0.33	0.66	0.83	0.09	0.14	0.18	0.28	0.57	0.04

Table 5 showed the association results between cognitive scores and the top 5 SNPs associated with T2w neuroimaging measurements based on distance covariance test. For each SNP, calculate the p-value on each cognitive score via a linear regression without (1st row) or with (2nd row) T2w measurements and GAF, age, gender, and socioeconomic status (house hold income and the highest parental education) as covariates. See Table 2 for exam descriptions.