Bivariate Genome-Wide Association Study of Genetically Correlated Neuroimaging Phenotypes from DTI and MRI through a Seemingly Unrelated Regression Model

Neda Jahanshad[1,**], Priya Bhatt[1,**], Derrek P. Hibar[1], Julio E. Villalon[1], Talia M. Nir[1], Arthur W. Toga[1], Clifford R. Jack, Jr.[2], Matt A. Bernstein[2], Michael W. Weiner[3,4], Katie L. McMahon[5], Greig I. de Zubicaray[6], Nicholas G. Martin[7], Margaret J. Wright[7], and Paul M. Thompson[1,*]

[1] Imaging Genetics Center, Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Los Angeles, CA
[2] Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA
[3] Department of Radiology, Medicine, and Psychiatry, UC San Francisco, CA, USA
[4] Department of Veterans Affairs Medical Center, San Francisco, CA, USA
[5] University of Queensland, Centre for Advanced Imaging, Brisbane, Australia
[6] University of Queensland, School of Psychology, Brisbane, Australia
[7] Queensland Institute of Medical Research, Brisbane, Australia

Abstract. Large multisite efforts (e.g., the ENIGMA Consortium), have shown that neuroimaging traits including tract integrity (from DTI fractional anisotropy, FA) and subcortical volumes (from T1-weighted scans) are highly heritable and promising phenotypes for discovering genetic variants associated with brain structure. However, genetic correlations (r_g) among measures from these different modalities for mapping the human genome to the brain remain unknown. Discovering these correlations can help map genetic and neuroanatomical pathways implicated in development and inherited risk for disease. We use structural equation models and a twin design to find r_g between pairs of phenotypes extracted from DTI and MRI scans. When controlling for intracranial volume, the caudate as well as related measures from the limbic system - hippocampal volume - showed high r_g with the cingulum FA. Using an unrelated sample and a Seemingly Unrelated Regression model for bivariate analysis of this connection, we show that a multivariate GWAS approach may be more promising for genetic discovery than a univariate approach applied to each trait separately.

Keywords: Neuroimaging genetics, brain connectivity, bivariate analysis, GWAS, genetic correlation.

[*] The Alzheimer's Disease Neuroimaging Initiative (ADNI).
[**] Equally contributing authors.

L. Shen et al. (Eds.): MBIA 2013, LNCS 8159, pp. 189–201, 2013.

1 Introduction

In many case-control genome-wide association studies (GWAS) of neurological diseases and disorders, such as Alzheimer's disease and schizophrenia, single genetic variations in the genome have been found to associate with the presence of the disease. While extremely popular, genetic association studies do not reveal what causes the disease or the neurological systems affected by the risk genes. On the other hand, if quantitative traits derived from brain imaging measures are used as a target for such genetic studies (e.g., GWAS), specific genetic loci can be mapped to specific features of the brain that may modulate disease risk.

Even so, single-nucleotide variations generally have small effects on any phenotype – carriers of different genetic variants may differ in the mean value of a brain measure by as little as 1-2%, or even less. Because of this, large populations are needed for genome-wide association analyses – making it very difficult to collect enough data at a single site to detect the effects on the brain of single-letter changes in the genome. Efforts to boost power in imaging genetics studies are therefore highly beneficial. Consortia that bring together information from multiple sites, such as the ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis) Project (http://enigma.loni.ucla.edu/), draw on information from tens of thousands of subjects [1] to boost the power to discover genetic influences on brain structure. These consortia require that the phenotype of interest be easily measurable, stable, robust to imaging acquisition parameters, and heritable across multiple cohorts. These studies often focus on bilaterally averaged measures of regional brain volumes with high signal-to-noise ratios, limiting the scope of possible phenotypes to those obtainable rapidly and routinely in many image analysis labs [2]. Despite these constraints on the kinds of brain measures that can be assessed in vast samples, several phenotypes from T1-weighted structural scans, and fractional anisotropy maps from diffusion tensor imaging scans, have been found to be feasible targets for large-scale genetic analysis. Large scale GWAS analyses have been applied to individual measures from brain MRI and DTI, analyzed one at a time, such as average bilateral subcortical volumes, and averaged bilateral mean FA values in tract-based regions of interest (ROIs).

Genetic epidemiological studies show that bivariate rather than univariate methods can improve power to identify and localize chromosomal regions associated with genetically correlated traits (in the case of linkage analysis, for example) [3]. Intuitively, if the same genetic variants influence two or more different imaging measures – even measures from different imaging modalities – it should be possible to measure the overlap and use it to boost the power to find which specific genetic variants are associated with each measure. This "genetic correlation" principle has been used recently to boost power in a variety of genetic analyses. It is especially valuable in imaging genetics because of the extreme difficulty of amassing samples large enough to pick up the effects of single genetic variants on the brain.

Even within an imaging modality such as DTI, Chiang et al. [4] recently used cross-twin cross-trait analysis to find sets of voxels in the brain that had common genetic influences. By clustering these voxels according to their genetic correlation, it was possible to screen the genome for variants that affected different regions of the image,

yielding higher power than simply considering each voxel on its own. The use of genetic correlation to discover coherent patterns of genetic influence in an image can also be extended to pick up common genetic influences on data from different imaging modalities. Recently, classical methods from quantitative genetics have been used to reveal bivariate genetic correlations among brain measures from T1-weighted MRI and DTI [5] - the two modalities used in this paper. The prior study by Kochunov et al., proved successful in using a large pedigree sample to find genetic correlations between brain measures from two imaging modalities, and localizing loci on the genome that had suggestive associations to phenotypes from both modalities, including average global FA and cortical thickness. Additionally, these prior works used genetic correlations to boost power to find specific genetic variants of interest. We set out to expand this type of analysis to multiple regions and phenotypes extractable in large cohorts and consortia such as ENIGMA. The overarching goal of our work is to discover genetic variants that affect brain measures that can be extracted efficiently from large neuroimaging databases worldwide. One tactic to do so, as shown in this paper, is to probe multiple imaging modalities for common genetic influences, and use these patterns as a coherent target to hunt for influential genes.

The Seemingly Unrelated Regression (SUR; [6]) bivariate model has proven to be a successful model to evaluate simultaneous GWASs of correlated traits. Unlike other multivariate methods, the SUR model is a system of linear equations that enables an unrestricted bivariate association test of genetic effects on each trait. In addition, the SUR model provides a great deal of flexibility as it includes a set of simultaneous regression equations that can have different sets of dependent variables and predictors. SUR can be thought of as a generalization of multiple regression, in which several multiple regression equations are all satisfied at once. The idea behind SUR is to fit a number of regression equations at once – not necessarily with the same outcome measure – and use the information on the covariance in the errors from each equation to update the others. Here we aimed to study genes affecting brain measures computed from T1-weighted MRI and DTI scans. It is logical that the size and white matter integrity of different parts of the brain might share some degree of genetic correlation, and we set out to find these patterns. We studied a cohort of healthy young twins to determine the genetic correlations between the different MRI and DTI measures. We then followed through with a bivariate SUR genome-wide association analysis of some of the most correlated regions in an unrelated elderly cohort (ADNI) with various degrees of cognitive impairment, including the FA of the cingulum and the volume of the caudate and the hippocampus.

2 Methods

2.1 Image Acquisition and Subject Information

We analyzed MRI and DTI data from two separate cohorts of human subjects: QTIM and ADNI; the details of these cohorts and the methods used to scan them are summarized below. QTIM, a large dataset of twins imaged with both MRI and high

angular resolution diffusion imaging, was used to assess the genetic correlations between phenotypes from the different modalities using a bivariate twin modeling design. ADNI, a publicly available dataset that includes subject genotypes, neuroimaging scans (MRI, DTI, rsfMRI, etc.), and a whole host of biological and clinical assessments, was used for the genome-wide association study.

QTIM (Queensland Twin IMaging study) -- Whole-brain 3D anatomical MRI and HARDI scans were acquired from over 600 healthy genotyped twins and siblings (age: 23.1±2.0) with a high magnetic field (4T) Bruker Medspec MRI scanner. T1-weighted images were acquired with an inversion recovery rapid gradient echo sequence. Acquisition parameters were: TI/TR/TE=700/1500/3.35 ms; flip angle=8 degrees; slice thickness = 0.9mm, with a 256x256 acquisition matrix. Diffusion-weighted images (DWI) were acquired using single-shot echo planar imaging with a twice-refocused spin echo sequence to reduce eddy-current induced distortions. Imaging parameters were: 23 cm FOV, TR/TE 6090/91.7 ms, with a 128x128 acquisition matrix. Each 3D volume consisted of 55 2-mm thick axial slices with no gap and 1.79x1.79 mm^2 in-plane resolution. 105 images were acquired per subject: 11 with no diffusion sensitization (i.e., b_0 images) and 94 DWI (b=1159 s/mm^2) with gradient directions evenly distributed on the hemisphere. Scan time was 14.2 minutes. In total, images from 224 of the entire group of right-handed young adults (mean age: 23.4 years, s.d. 2.0) were analyzed in this study, including 51 pairs of monozygotic twins and 56 pairs of dizygotic twins. Remaining subjects included non-twin siblings, singletons, and those subjects whose image processing and volume extraction did not pass rigorous quality control. This set of data comprised the Queensland Twin Imaging dataset, QTIM.

ADNI -- The Alzheimer's Disease Neuroimaging Initiative (ADNI), is a multisite longitudinal study comprised of clinical, genetic and neuroimaging data of AD, MCI and normal elderly patients with varying degrees of cognitive impairment. ADNI recently launched a second phase (ADNI-2) of longitudinal data collection to include diffusion-weighted scans, among other newly added imaging modalities, with the goal of studying microstructural integrity and anatomical connectivity (among other measures) in elderly individuals. Whole-brain MRI scans were collected using 3-Tesla GE Medical Systems scanners, at 14 sites across North America. T1-weighted SPGR (spoiled gradient echo) anatomical scans were collected in addition to DWIs. For each diffusion MRI scan, there were 41 DWIs (b=1000 s/mm^2) and 5 T2-weighted images acquired with no diffusion gradient (b_0 images). This protocol was selected to optimize the signal-to-noise ratio given a fixed scan time [7]. At the time of writing (June 2013), approximately 200 subjects have been scanned with DTI. Of those scanned so far with DTI, 65 of them possessed sufficient genetic data and imaging data, including data from both DTI and MRI scans. A blood draw for genomic DNA extraction of each subject was obtained at the screening or baseline

visit and was genotyped using the Illumina HumanOmniExpress BeadChip (Illumina, Inc, San Diego, CA, USA) for each subject (620,901 SNPs). Any SNP with a minor allele frequency less than 0.1 was excluded from the study. Table 1 shows a summary of relevant demographic information for this study.

Table 1. The highest degree of genetic correlation was shared between the FA of the cingulum and the volume of the caudate nucleus of the brain. Additionally the cingulum mean FA was also the tract measure most highly genetically correlated with the volume of the hippocampus, both of which make up the circuitry of the limbic system – a key target of pathology in the development of Alzheimer's disease. Average volume, FA, and standard deviation of structures are shown.

	Caudate Volume Mean (SD)	Hippocampal Volume Mean (SD)	Cingulum FA Mean (SD)	Age (yrs) Mean (SD)	ICV
Total n = 65	3288.5 mm^3 (481.8)	3367.9 mm^3 (497.7)	0.362 (0.04)	74.42 (7.4)	0.765 (0.09)
Males n = 36	3455.5 mm^3 (500.6)	3460.8 mm^3 (550.2)	0.370 (0.04)	73.57 (7.1)	0.817 (0.06)
Females n = 29	3081.1 mm^3 (369.9)	3252.5 mm^3 (403.8)	0.351 (0.03)	75.48 (7.8)	0.701 (0.06)

2.2 DTI and MRI Processing and Phenotype Extraction

Non-brain regions were automatically removed from each T1-weighted MRI scan, and from a T2-weighted image from the DWI set using the FSL tool "BET" (http://fsl.fmrib.ox.ac.uk/fsl/). A trained neuroanatomical expert manually edited the T1-weighted scans to further refine the brain extraction. All T1-weighted images were linearly aligned using FSL (with 9 DOF) to a common space with 1mm isotropic voxels and a 220×220×220 voxel matrix. DWI data were corrected for eddy current distortions using the FSL tools (http://fsl.fmrib.ox.ac.uk/fsl/). For each subject, the images with no diffusion sensitization were averaged, and elastically registered to the T1 scan to compensate for susceptibility artifacts.

For both datasets, single tensor-based FA maps were computed from the DW images and registered to the ENIGMA-DTI mean template [2, 8] as outlined at http://enigma.loni.ucla.edu/ongoing/dti-working-group/. Skeletonized maps were created using tract-based spatial statistics (TBSS [9]) and regions along the skeleton were parcellated according to the Johns Hopkins University DTI Atlas [10]. The FA of partial tracts and bilateral regions were averaged. The regions analyzed (ROIs) included the full skeleton, GCC – the *genu* of the corpus callosum, BCC – the body of the corpus callosum, SCC – the *splenium* of the corpus callosum, FX – the fornix, CGC – the bilateral cingulate, CR – *corona radiata*, EC – bilateral external capsule, IFO – bilateral inferior fronto-occipital fasciculus, PTR – posterior thalamic radiation, SFO – bilateral superior fronto-occipital fasciculus, SLF – bilateral superior longitudinal fasciculus, SS – bilateral sagittal stratum, and CST – the bilateral corticospinal tract.

Subcortical regions were segmented using FSL's FIRST according to protocols found at http://enigma.loni.ucla.edu/protocols/imaging-protocols/ [11]. These regions included average bilateral volumes for the amygdala, nucleus accumbens, hippocampus, caudate, thalamus, pallidum, and putamen.

2.3 Calculating the Genetic Correlation

Using the QTIM sample, we used a "cross-twin cross-trait" analysis to detect common genetic or environmental factors influencing all pairs of subcortical structure volumes and the mean FA for all regions of interest examined. Covariance matrices for these phenotypes were computed for the monozygotic twins (MZ) who share all the same genes, and the dizygotic twins (DZ) who share on average half of their genetic polymorphisms. These covariance matrices were then entered into a multivariate structural equation model (SEM) using OpenMx [12] (http://openmx.psyc.virginia.edu) to fit the relative contributions of additive genetic (A), shared environmental (C), and unshared or unique environmental and error (E) components to the population variances and covariances of the observed variables.

If the correlation between the volume in one twin and the FA in the other twin is greater in MZ pairs (sharing 100% of their genome) than in DZ pairs (sharing approximately 50% of their genome), then, in the classical twin design, we assume that the greater correlation may be attributable to common genetic factors that influence the two traits. Using a multivariate SEM, we can compute the additive genetic and shared environmental influences on the correlations between the two phenotypes, denoted as r_A and r_C, respectively. Significant associations were determined if removal of the additive genetic correlation component of the model (r_A in the path diagram seen in Figure 1) resulted in a significantly poorer fit to the data. Using this model, we were able to narrow down regions where a strong bivariate genetic association was detectable. This would imply that both the FA and the volumes of the respective brain regions are influenced by a common subset of additive genetic factors, or that these two imaging traits exhibit "pleiotropy". Pleiotropic effects of genes in this case would imply that a trait in twin #1 of a pair is able to predict the other trait in the twin #2 of the pair, and that the predictions tend to be better in identical than fraternal twins; if common genes are responsible for driving some of the correlation, the cross-twin prediction would be more accurate in the case of MZ twins than DZ twins as they share more of their genome. With a large enough sample size, recent methods show that it may be possible to estimate the shared genetic correlation of different traits using the full genome even in unrelated individuals [13].

2.4 Seemingly Unrelated Regression Model for GWAS

A standard univariate genome-wide association test uses the general linear regression model to test the additive effect of all genotyped variants on the phenotype of interest.

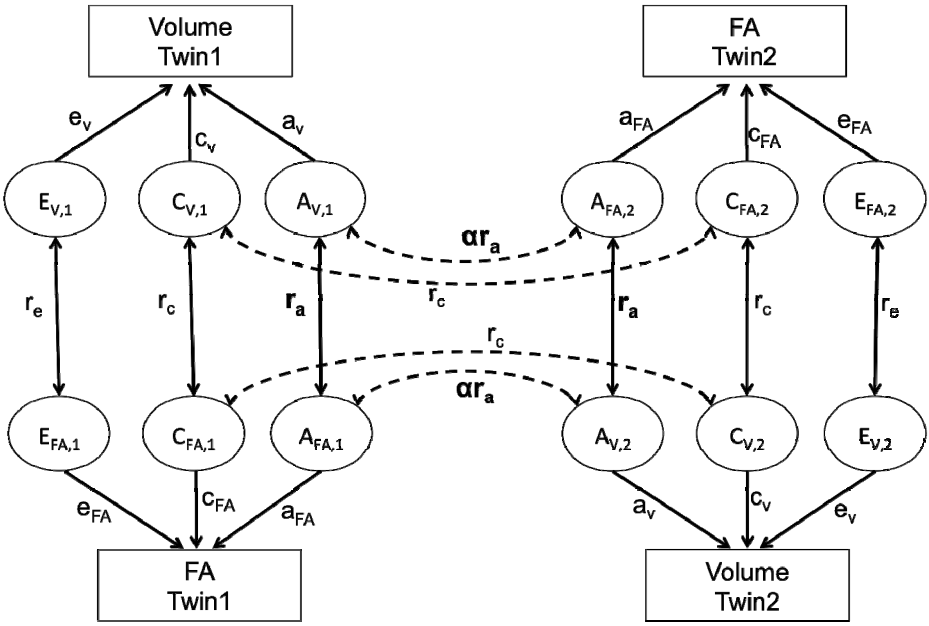

Fig. 1. Cross-twin cross-trait diagram for computing genetic correlations. In this method from classical quantitative genetics, a trait in one twin (such as the volume of part of the brain) is used to predict a different trait in the other twin (such as the integrity of a fiber tract on DTI). If these "cross-twin cross-trait" predictions can be made more accurately in identical twins, who share all their genes, than in fraternal twins, who do not, there is a basis to begin to model the additive genetic influences on both traits, as well as many other components that account for the correlations between the imaging modalities.

Using ADNI, a neuroimaging dataset independent of the one we used to calculate genetic correlations, we ran a univariate genome-wide association study (GWAS) on subcortical volume measures (the caudate and the hippocampus), which were found to have highest genetic correlations with the FA of the cingulum (see Results). As we searched the full genome, i.e., we tested approximately 500,000 SNPs (with a minor allele frequency (MAF) greater than 0.1) for any statistical association with each volume of interest; a stringent threshold is set for determining significance (as is standard practice in a GWAS study). We selected all variants that showed any sign of suggestive associations ($p < 10^{-4}$) to test in our SUR model. We hypothesized that some of these variants are also associated with the FA of the cingulum; we expected that joint bivariate analysis would allow an improvement in the significance of the tested associations.

The seemingly unrelated regression (SUR) model was originally developed for applications in econometrics [6]. Recently, the SUR model has been applied in a biomedical framework, i.e., identifying genetic variants associated with

endophenotypic traits of Alzheimer's Disease [14] and bone marrow density [15] as well as identifying connectivity failures related to cognitive decline in Alzheimer's Disease [16]. The SUR model is a system of linear regression models for a number of traits, n (in our case we have two, y_1 and y_2). It assumes that the residual error terms (ε) are identically and independently distributed for each individual (1, …, N) within traits. The idea behind the model is to fit a number of regression equations at once – not necessarily with the same outcome measure, and use the information on the covariance in the errors from each equation to update the other. These linear regression models are "related" by their correlated residual error terms. In addition, the SUR model provides flexibility by allowing a different set of explanatory variables (X_i) to predict different traits, with the idea that some variables may be associated with only one trait. The SUR model can be written as:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

(Equation 1)

SUR allows us to fit a number of regression equations at once, using the solution of a set of simultaneous equations to improve the accuracy of all the fitted models. For the purposes of this study, we used a bivariate SUR model of correlated brain measures to help boost the power to detect those genetic associations often too difficult to discover due to their small effect sizes. The two traits examined are a given subcortical volume (e.g., the volume of the hippocampus, averaged across the left and right hemispheres) and an average tract FA (e.g., the mean FA of the cingulum). Specifically for the purposes of this study, we examined two bivariate models: (1) the average volume of the hippocampus and the average tract FA and (2) the average volume of the caudate and the average FA tract. In our regression models that are assumed to hold simultaneously, the regional brain volumes are predicted by sex, age and intracranial volume and FA is predicted by sex and age, where coefficients for all variables remained unrestricted. Additionally (not shown in the equation) we control for the first four components derived from the multi-dimensional scaling (MDS) plots to control for differences in population structure, an important factor in genetic association studies. This correction makes sure that differences in ethnicity among participants do not lead to spurious genetic associations with brain measures of interest that would not hold up within any one ethnic group.

$$\begin{matrix} y_{hvol} = Age * \beta_{age} + Sex * \beta_{sex} + IVC * \beta_{IVC} + a_{hvol} + \varepsilon_{hvol} \\ y_{CGC} = Age * \beta_{age} + Sex * \beta_{sex} + a_{CGC} + \varepsilon_{CGC} \end{matrix} + \varepsilon$$

(Equation 2)

SUR regressions were carried out using the 'systemfit' package in R (version 2.15.1).

3 Results

The cross-twin cross-trait model was fitted between all pairs of bilateral volume measures and FA measures. When controlling for the intracranial volume, this model revealed a high genetic correlation between the average caudate volume and the mean FA in the bilateral cingulum (Figure 2). This relationship is of interest, as it is not obvious *a priori* that measures from DTI will correlate with morphometry on standard anatomical MRI, and it is even less obvious that the correlations will be due to common variants in the genome. The presence of genetic correlations across imaging modalities shows that both of these situations are indeed true of our datasets.

Additionally, the FA of the cingulum is the connection most genetically correlated with the hippocampus - a region previously used as the phenotype of interest to discover novel genetic associations in the largest meta-analytical brain imaging studies to date (N>20,000 subjects) [1, 17]. Although the genetic correlation is much lower, it is still significant.

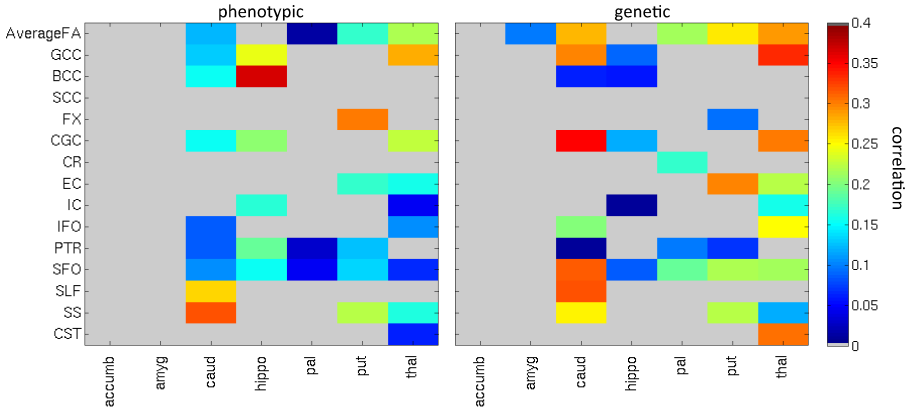

Fig. 2. Left: phenotypic correlations between different brain measures, and **Right:** genetic correlations between bilateral subcortical structure volumes (from anatomical MRI) and mean tract FAs (from DTI scans of the same subjects) while controlling for the intracranial volume, where *red colors* indicate high genetic correlation and *blue* indicates low detectable correlations. Note that the phenotypic correlation is the standard correlation between any of the two brain measures, here taking into account the non-independence of siblings by using a random-effects model. The genetic correlation will be zero if the correlation is zero, or if it is non-zero but there are no genetic variants with detectable effects on both measures.

A univariate GWAS was performed for each of the caudate and the hippocampal volumes. We controlled for age, sex, intracranial volume, and the first four components of the multidimensional scaling (MDS) plots. All SNPs that showed a suggestive association ($p<10^{-4}$) were carried forward into an SUR model. This included 39 SNPs for the caudate and 30 SNPs for hippocampal volume. The bivariate SUR model involved the volume of interest and also the FA of the cingulum, which was the DTI measure found to have the highest genetic correlation with both

volumes. In Figure 3, results from both univariate and bivariate analyses are plotted relative to the (uniform) *p*-values that would be expected under the null hypothesis. In general, while it appears the bivariate assessment does not improve association *p*-values for the most associated variants, there is a general trend for improving the power for association overall.

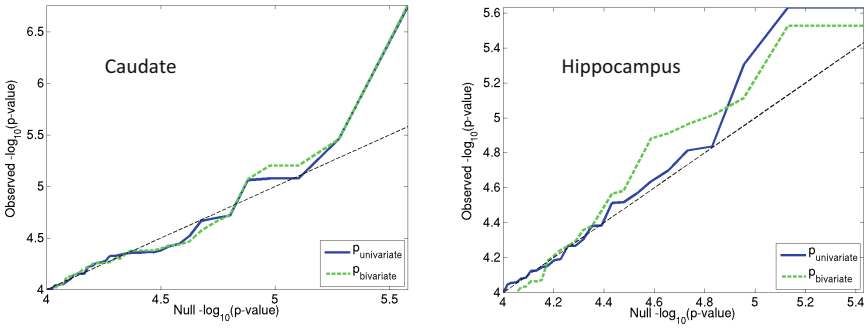

Fig. 3. All genetic variants that showed a suggestive statistical association ($p < 10^{-4}$) with (a) caudate nucleus volume, or (b) hippocampal volume, when controlling for age, sex, intracranial volume and the first four MDS components, were carried forward into an SUR model using the effect of the variant on the genetically correlated cingulum FA measure as a second predictive equation, controlling for age, sex, and the first four MDS components. Results from both univariate and bivariate analyses are plotted relative to the (uniform) *p*-values that would be expected under the null hypothesis. In general, while it appears the bivariate assessment does not improve association *p*-values for the most associated variants, there may be a general trend for improving the power for association overall. This is just a proof of concept study, and expansion of the methods to a larger cohort may corroborate the theoretical advantage of bivariate correlation methods in boosting the power of GWAS.

4 Discussion

Here we developed a new method to find genetic factors that affect two different brain imaging modalities, after showing that measures from each are genetically correlated. This leads to the remarkable conclusion that common variants in the human genome are partly responsible for driving the correlation between measures in two very different imaging modalities. To show this correlation, we first determined the genetic correlation between various neuroimaging measures obtained from DTI and MRI, using a structural equation model and a twin design. Next, we show that these genetically correlated measures can be used in a bivariate genome-wide association test. This may, at least in principle, provide boosted power in identifying additional genetic loci involved in determining and influencing brain structure.

Using genetic correlations to cluster imaging phenotypes has previously been helpful in discovering brain regions influenced by the same underlying genetic

variants. One study clustered regions of the cortex with common underlying genetic determination [18,19]. It is logical that after clustering, this new regions of interest can be useful for boosting the power of univariate genome-wide association studies to discover genetic variants that may modulate brain structure using measures obtained from a single modality including DTI [4] or details of subcortical structures segmented from standard MRI [20]. Here, the joint use of the DTI and MRI measures offers additional advantages, as they are already measures known to be reliable and heritable [2, 21]. They are also the target biomarkers and phenotypes of several multisite projects and consortia, particularly of the ENIGMA group [22]. In such consortia, meta-analysis of genetic effects can discover and validate the effect of specific genetic variants on specific brain structures. These discovered variants may hold the key to identifying new genetic markers that put the brain at risk for developmental and degenerative disorders, while also allowing us to identify the neuroanatomical and molecular pathways involved.

A bivariate analysis such as the one proposed here – if conducted on a multi-site, meta-analytic scale – could boost power to discover genetic associations, and could help to identify molecular pathways in which these genes exert their effects. The same approach of clustering genetically correlated features in images could allow us to map the trajectory of effects on connected brain regions or even hubs and subnetworks in brain connectivity maps. Our SUR model even allows us to use different predictive covariates for different phenotypes of interest (i.e., using ICV as a covariate for volume measures, but not for FA measures from DTI scans).

ADNI, in particular, focuses on discovering genetic markers associated with brain degeneration in Alzheimer's disease. As such, the genetic correlations between measures from the cingulum in the limbic system and the hippocampus is of interest. The cingulum is the main white matter pathway connecting limbic structures (i.e., hippocampus, amygdala) with the cingulate cortex and the striatum. Our test set was comprised of only 65 unrelated subjects, so power is obviously extremely limited. While no genome-wide significant results are expected, our methods could be used to pick up trends for better identifying loci of interest. As the ADNI-2 DTI database grows, a future analysis involving more subjects is feasible.

A related idea to that in this paper is the use of very large sets of phenotypes to fit models, without requiring any of the subjects to have all the measures collected. Using imaging data from multiple sources, [23] were able to classify ADNI subjects into diagnostic groups, even when very large amounts of data are missing. The use of sparse regression models that allow "block-wise" missing data is a considerable advantage in epidemiology, as there are many cohorts with "deep phenotyping" – many measures assessed – without being able to rely on having all measures in all subjects. When the available sample sizes become truly vast (such as the 10,000+ subjects analyzed in the ENIGMA efforts [22,24-26]), it should be possible to fit bivariate correlation matrices between all pairs of points in a pair of imaging modalities, followed by clustering of the resulting genetic correlation matrix. Such an approach has already been found to boost power for GWAS studies within a modality [17], but the current paper suggests that it is possible to draw upon a very large body

of multi-modal biomarker data and apply genetic screening to pick out coherent inherited patterns that run through all imaging modalities. Among all the ways to boost power in imaging genetics, it seems that the joint and simultaneous use of many modalities of data may help pick up genetic signals. In this way, imaging genetics offers some advantages over clinical genetic association studies, as the target signal is a vast and high-dimensional signal with latent structure. The better we exploit the underlying statistical coherence in the target signals, the more rapidly and efficiently we will detect the genetic variants that influence them.

References

1. Stein, J.L., et al.: Identification of common variants associated with human hippocampal and intracranial volumes. Nat. Genet. 44(5), 552–561 (2012)
2. Jahanshad, N., et al.: Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA-DTI working group. Neuroimage (2013)
3. Almasy, L., Dyer, T.D., Blangero, J.: Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. Genetic Epidemiology 14(6), 953–958 (1997)
4. Chiang, M.C., et al.: Gene network effects on brain microstructure and intellectual performance identified in 472 twins. J. Neurosci. 32(25), 8732–8745 (2012)
5. Kochunov, P., et al.: Genetic analysis of cortical thickness and fractional anisotropy of water diffusion in the brain. Frontiers in Neuroscience 5, 120 (2011)
6. Zellner, A.: An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. Journal of the American Statistical Association 57(298), 348–368 (1962)
7. Zhan, L., et al.: Angular versus spatial resolution trade-offs for diffusion imaging under time constraints. Hum. Brain Mapp. (2012)
8. Kochunov, P., et al.: Genome-wide association of full brain white matter integrity – from the ENIGMA DTI working group. In: Organization of Human Brain Mapping, Beijing, China (2012)
9. Smith, S.M., et al.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage 31(4), 1487–1505 (2006)
10. Mori, S., et al.: Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. Neuroimage 40(2), 570–582 (2008)
11. Hibar, D.P., +200-co-authors, ENIGMA-Consortium: ENIGMA2: Genome-wide scans of subcortical brain volumes in 16,125 subjects from 28 cohorts worldwide. In: Organization of Human Brain Mapping, Seattle, WA (2013)
12. Boker, S., et al.: OpenMx: An Open Source Extended Structural Equation Modeling Framework. Psychometrika 76(2), 306–317 (2011)
13. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M., Wray, N.R.: Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 28, 2540–2542 (2012)
14. Bhatt, P., et al.: Multivariate analysis of GWAS for identification for genetic variants in Endophenotypes related to Alzheimer's Disease, Master's thesis. Oregon Health and Science University (2012)

15. Saint-Pierre, A., et al.: Bivariate association analysis in selected samples: application to a GWAS of two bone mineral density phenotypes in males with high or low BMD. Eur. J. Hum. Genet. 19(6), 710–716 (2011)
16. Jahanshad, N., et al.: Boosting power to associate brain connectivity measures and dementia severity using Seemingly Unrelated Regressions (SUR). In: Wang, L., Yushkevich, P., Ourselin, S. (eds.) MICCAI Workshop on Novel Imaging Biomarkers in Alzheimer's Disease, Nice, France. LNCS, pp. 103–112 (2012)
17. Bis, J.C., et al.: Common variants at 12q14 and 12q24 are associated with hippocampal volume. Nat. Genet. 44(5), 545–551 (2012)
18. Chen, C.H., et al.: Hierarchical genetic organization of human cortical surface area. Science 335, 1634–1636 (2012)
19. Chen, C.H., et al.: Genetic influences on cortical regionalization in the human brain. Neuron 72, 537–544 (2011)
20. Hibar, D.P., et al.: Genetic clustering on the hippocampal surface for genome-wide association studies. In: Mori, K., et al. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 674–681. Springer, Heidelberg (2013)
21. den Braber, A., Bohlken, M.M., Brouwer, R.M., van 't Ent, D., Kanai, R., Kahn, R.S., de Geus, E.J., Hulshoff Pol, H.E., Boomsma, D.I.: Heritability of subcortical brain measures: A perspective for future genome-wide association studies. Neuroimage (2013)
22. Thompson, P.M., et al.: The ENIGMA Consortium: Large-scale Collaborative Analyses of Neuroimaging and Genetic Data. Special Issue of Brain Imaging and Behavior, Invited Review (in submission, 2013)
23. Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J.: Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. Neuroimage 61, 622–632 (2012)
24. Hibar, D.P., et al., ENIGMA-Consortium: ENIGMA2: Genome-wide scans of subcortical brain volumes in 16,125 subjects from 28 cohorts worldwide. In: Organization of Human Brain Mapping, Seattle, WA (2013)
25. Turner, J.A., et al., ENIGMA-Schizophrenia: A Prospective Meta-Analysis of Subcortical Brain Volumes in Schizophrenia via the ENIGMA Consortium. In: Organization of Human Brain Mapping, Seattle, WA (2013)
26. Hibar, D.P., et al., ENIGMA-BipolarDisorder: Meta-analysis of structural brain differences in bipolar disorder: the ENIGMA-Bipolar Disorder. In: Organization of Human Brain Mapping, Seattle, WA (2013)