

Published in final edited form as:

*Neuroimage*. 2011 June 15; 56(4): 1875–1891. doi:10.1016/j.neuroimage.2011.03.077.

## Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects

Derrek P. Hibar<sup>1</sup>, Jason L. Stein<sup>1</sup>, Omid Kohannim<sup>1</sup>, Neda Jahanshad<sup>1</sup>, Andrew J. Saykin<sup>2</sup>, Li Shen<sup>2</sup>, Sungeun Kim<sup>2</sup>, Nathan Pankratz<sup>3</sup>, Tatiana Foroud<sup>3</sup>, Matthew J. Huentelman<sup>4</sup>, Steven G. Potkin<sup>5</sup>, Clifford R. Jack Jr.<sup>6</sup>, Michael W. Weiner<sup>7,8</sup>, Arthur W. Toga<sup>1</sup>, Paul M. Thompson<sup>1</sup>, and the Alzheimer's Disease Neuroimaging Initiative\*

<sup>1</sup> Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Los Angeles, CA

<sup>2</sup> Center for Neuroimaging, Department of Radiology and Imaging Science, Indiana University School of Medicine, Indianapolis, IN

<sup>3</sup> Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN

<sup>4</sup> The Translational Genomics Research Institute, Phoenix, AZ

<sup>5</sup> Department of Psychiatry and Human Behavior, University of California, Irvine, Irvine, CA

<sup>6</sup> Mayo Clinic, Rochester, MN

<sup>7</sup> Departments of Radiology, Medicine, Psychiatry, UC San Francisco, San Francisco, CA

<sup>8</sup> Department of Veterans Affairs Medical Center, San Francisco, CA

### Abstract

Imaging traits provide a powerful and biologically relevant substrate to examine the influence of genetics on the brain. Interest in genome-wide, brain-wide search for influential genetic variants is growing, but has mainly focused on univariate, SNP-based association tests. Moving to gene-based multivariate statistics, we can test the combined effect of multiple genetic variants in a single test statistic. Multivariate models can reduce the number of statistical tests in gene-wide or genome-wide scans and may discover gene effects undetectable with SNP-based methods. Here we present a gene-based method for associating the joint effect of single nucleotide polymorphisms (SNPs) in 18,044 genes across 31,662 voxels of the whole brain in 731 elderly subjects (mean age:  $75.56 \pm 6.82$ SD years; 430 males) from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Structural MRI scans were analyzed using tensor-based

© 2011 Elsevier Inc. All rights reserved.

Please address correspondence to: Paul Thompson PhD, Professor of Neurology, Laboratory of Neuro Imaging, Dept. of Neurology, UCLA School of Medicine, Neuroscience Research Building 225E, 635 Charles Young Drive, Los Angeles, CA 90095-1769, USA, Phone: (310) 206-2101 Fax: (310) 206-5518 thompson@loni.ucla.edu.

\*Data used to prepare this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([www.loni.ucla.edu/ADNI](http://www.loni.ucla.edu/ADNI)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. For a complete listing of ADNI investigators, please see: [http://www.loni.ucla.edu/ADNI/Collaboration/ADNI\\_Manuscript\\_Citations.pdf](http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf)

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

morphometry (TBM) to compute 3D maps of regional brain volume differences compared to an average template image based on healthy elderly subjects. Using the voxel-level volume difference values as the phenotype, we selected the most significantly associated gene (out of 18,044) at each voxel across the brain. No genes identified were significant after correction for multiple comparisons, but several known candidates were re-identified, as were other genes highly relevant to brain function. *GAB2*, which has been previously associated with late-onset AD, was identified as the top gene in this study, suggesting the validity of the approach. This multivariate, gene-based voxelwise association study offers a novel framework to detect genetic influences on the brain.

## Keywords

principal components regression; voxelwise; multivariate; gene-based; GWAS; *GAB2* (max. 6 keywords)

## 1. Introduction

Recent efforts in imaging genetics have advanced the field rapidly from identifying heritable features of the brain to genome-wide searches for specific genetic variants that might account for functional and structural variations in large populations (Potkin et al., 2009a; Potkin et al., 2009b; Shen et al., 2010; Stein et al., 2010b; Thompson et al., 2010). Variation in the human genome may account for variations in brain integrity, and multi-national consortia have been set up to discover and verify genetic effects on brain images (e.g., the ENIGMA project; <http://enigma.ion.ucla.edu>). In imaging genomics, the vast amount of information in the images (>100,000 voxels) and across the genome (>12 million known variants) requires powerful methods to relate genetic variants to the structure and function of the brain. Power issues arise due to the small effect sizes, and the huge numbers of statistical comparisons. Most techniques use some type of data reduction, limiting the number of genetic variants studied or the number of imaging features studied, or both. The ultimate goal of these gene-hunting studies is to create a method that addresses the gene discovery problem in a statistically powerful and biologically meaningful way.

The current mainstay of gene-hunting efforts in imaging genetics is the genome-wide association study (*GWAS*). Most genetic association tests relate individual SNPs to phenotypes, but since there are, on average between 20 and 100 SNPs per gene (in our dataset), and alleles at these SNPs are often highly correlated, a method which tests all the SNPs in a gene at once (or most of the variance contributed by SNPs in a gene) would reduce the number of tests required and be more powerful. We will hereafter refer to SNP-Based approaches and gene-based approaches. This assesses associations between common SNPs and features in an image. In typical GWAS studies, each genetic variant (usually a SNP) is independently tested for its association to the phenotype – a mass univariate method, where no data reduction is used across the genome. For example, Stein et al. (2010b) performed a genome-wide search of around 500,000 SNPs, and found a novel variant in the *GRIN2B* gene that is associated with temporal lobe volume. The gene *GRIN2B* encodes a glutamate receptor that is already the target of drugs (memantine) used to treat Alzheimer's disease (Parsons et al., 2007). Findings such as these are promising as they have biological relevance without relying on a prior hypothesis about any specific SNP. However, performing mass SNP-based tests on imaging summary measures (such as temporal lobe volume, hippocampal volume, etc.) or *ad hoc* regions of interest (ROI), collapses the brain measures into a single number. Studies using an ROI to define the imaging phenotype may miss fine-grained differences throughout the brain, across subjects.

In addition, a predefined ROI can lead to false-negative results if a true association signal lies outside or only partially within a chosen ROI.

Several studies now perform genome-wide searches at each voxel across the brain (Hibar et al., 2010). This approach avoids pre-selecting an *ad hoc* region of interest in the brain and does not require prior hypotheses about which genetic variants, or which regions of interest, matter. Stein et al. (2010a) performed a genome-wide, brain-wide search, termed a voxelwise genome-wide association study (*vGWAS*), in 740 subjects from the ADNI. The experiment was extremely computationally intensive (27 hours on 500 nodes), performing around 16 trillion tests of association. However, the correction for multiple comparisons was commensurate with the number of tests performed. None of the variants identified was significant after multiple comparisons correction, but several variants were promising candidates for further analysis. In an alternative approach, Vounou et al. (2010) proposed a method that leverages the sparseness of signals to simultaneously select SNP variants and regions of association, reducing the number of SNPs and phenotypes tested. Future GWAS studies in imaging will likely reduce the number of tests and multiple comparisons using Bayesian priors. This can prioritize certain regions of the image or the genome, for later meta-analysis across multiple datasets.

*Gene-based* association methods complement single-marker GWAS for implicating underlying genetic variants in complex traits and diseases (Neale and Sham et al., 2004). Given recent advances in high-throughput genotyping, densely-packed sets of SNPs, or genetic markers, can capture increasing amounts of variation throughout the genome. Methods that consider combinations of SNPs from the same gene should better describe genetic associations than methods that rely on data from SNPs independently (Neale and Sham et al., 2004; Schaid et al., 2004). Whole-gene testing is a biologically plausible approach to the problem, as the ultimate unit of biological activity is the gene (or its protein product; Potkin et al., 2009c). By associating the joint effect of multiple SNPs within a gene, in this study we aimed to show that gene-based approaches can be more powerful than traditional SNP-based approaches (with the relative power depending on how the genetic variants affect the phenotype). For example, if a gene contains multiple causal variants with small individual effects, SNP-based methods will miss these associations if a very stringent significance threshold is used (as in GWAS). In addition, if multiple loci within a gene combine to jointly affect a phenotype, this may also be missed by traditional GWAS. These two scenarios are highly likely, especially if we accept the “common disease, common variant” hypothesis (Reich and Lander, 2001), but they are not accounted for in methods that test each SNP, one at a time.

A multi-SNP, gene-based test can consider the combined effect of each variant within the gene, while accounting for linkage disequilibrium (LD) or correlation between markers. As such, at least in theory it may detect associations missed by traditional SNP-based GWAS. Related to this approach is “multi-locus fitting” - a developing field in quantitative genetics, for the analysis of complex traits. Some multi-locus analyses use statistical methods specialized for handling high-dimensional data, including regularized regression methods such as ridge regression (Malo et al., 2008; Sun et al., 2009), the Bayesian lasso (Zou et al., 2006; Wu et al., 2009), and neural network models (Lucek et al., 1998; Ott et al., 2001). Another related approach is set-based association testing, implemented in the software Plink (Purcell et al., 2007), which allows for the combination of univariate test statistics into a single univariate test statistic using permutations. Gene-based tests also reduce the effective number of statistical tests by aggregating multiple SNP effects into a single test statistic. However, for gene-based tests to be feasible, the multivariate test statistics need to be computationally efficient to implement. Here we assessed whether it would be feasible to extend to a neuroimaging database, a gene-based association method using principal

components regression (PCReg) as proposed by Wang and Abbott (2008) for single-valued traits. We applied PCReg across all genes, to a large database of voxelwise imaging data. We call our method a voxelwise “gene-wide” association study (vGeneWAS). By performing association tests on whole genes, we greatly reduce the number of tests (from 437,607 SNPs down to 18,044 genes) while avoiding the problems associated with focusing on ROIs or summary measures. Our framework shows how to conduct vGeneWAS studies, and identify gene variants that warrant further study.

We hypothesized that vGeneWAS would, in some situations, have greater power to detect associations than existing SNP-based methods. One such situation might be when a gene contains many loci with weak individual effects. In addition, we expected that vGeneWAS would have greater overall power than mass SNP-based methods, like vGWAS, because of the drastic reduction in the effective number of statistical tests performed.

## 2. Materials and methods

### 2.1 Study design and subjects assessed

ADNI is a large 5-year study initiated in 2003 as a public-private partnership between the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations. The ADNI study aims to identify and investigate biological markers of Alzheimer’s disease through a combination of neuroimaging, genetics, neuropsychological tests and other measures in order to develop new treatments, track disease progression, and lessen the time required for clinical trials. The study was conducted according to the Good Clinical Practice guidelines, the Declaration of Helsinki, and U.S. 21 CFR Part 50—Protection of Human Subjects, and Part 56—Institutional Review Boards. Written informed consent was obtained from all participants before protocol-specific procedures were performed.

The study recruited 202 Alzheimer’s disease subjects (AD), 413 with mild cognitive impairment (MCI), and 237 normal elderly controls (NC) who were assessed every 6 or 12 months for three years. Subjects went through extensive clinical and cognitive tests at the time of each scan to determine and track diagnosis. Further information on inclusion criteria and the study protocol may be found online (<http://adni-info.org/>). Baseline structural MRI scans and genetic data for 818 subjects were obtained on or before May 5, 2010, from the public ADNI database (<http://www.loni.ucla.edu/ADNI/>). Scans for 852 subjects were available, but we excluded 121 subjects based on quality control measures (poor registration, image quality) and to avoid a well documented problem in statistical genetics known as population stratification (McCarthy et al., 2008). When performing association tests on latent subpopulations of different ethnicities or relatedness, spurious associations may arise due to differences in allele frequencies between groups, instead of true association with the phenotype (Lander and Schork, 2001). Subjects were removed based on self-reported ethnicity, later verified by multi-dimensional scaling (MDS) analysis (see previous study: Stein et al., 2010b), leaving 172 AD patients (78 women/94 men; mean age  $\pm$  standard deviation =  $75.54 \pm 7.62$  years), 356 MCI subjects (130 women/226 men; mean age:  $75.23 \pm 7.22$ ), and 203 healthy elderly controls (93 women/110 men; mean age:  $76.15 \pm 4.99$ ). We did not split the subjects by diagnosis for this analysis in order to exploit the broadest phenotypic continuum (Petersen et al., 2000) and maximize statistical power to detect genetic associations (Cannon and Keller, 2006).

### 2.2 Imaging methods

Baseline MRI scans for each subject were analyzed using tensor-based morphometry (TBM) as described previously (Hua et al., 2008). Briefly, high-resolution T1-weighted structural

brain MRI scans were acquired at 58 ADNI sites on 1.5T scanners with a protocol developed for multiple site consistency (Jack et al., 2008; ADNI also collected some data at 3T, which we did not analyze here; see Ho et al. 2010). Additional image corrections were applied to all images in a processing pipeline including: *GradWarp* correction of geometric distortion (Jovicich et al., 2006), B1-correction to adjust image intensity non-uniformity (Jack et al., 2008), N3 bias correction to adjust intensity inhomogeneity across a scan (Sled et al., 1998), and geometric scaling determined by a phantom scan acquired at each subject's scanning session to tune scanner and session-specific calibration errors (Jack et al., 2008). Images were linearly aligned using a 9 parameter algorithm to the International Consortium for Brain Imaging template (ICBM-53; Mazziotta et al., 2001) to align brain positions to a common standard space, adjusting for global scaling.

The TBM analysis was performed following the protocol of our prior study, which showed clinical and cognitive test scores correlated with temporal lobe volumes, in a subset of the ADNI population (Hua et al., 2008). A minimum deformation template (MDT) was created based on a random subset of the healthy elderly controls at baseline. The MDT provides an unbiased representation of MRI scans expected from a group of average healthy elderly persons. We generated maps of localized volume difference for each subject compared to the MDT using an inverse-consistent, symmetric, intensity-based nonlinear warping algorithm (Leow et al., 2006). Maps of localized volume differences (called Jacobian maps) are estimated using the Jacobian determinant of the deformation matrix, which itself is a voxel-level estimate of volume excess or deficit compared to the MDT. Jacobian maps for each individual were then down-sampled using trilinear interpolation to a  $4 \times 4 \times 4 \text{ mm}^3$  voxel size to reduce computational burden. The value at each voxel in the Jacobian map represents a percentage volume difference compared to the MDT; we used this voxel-based measure of volume difference as the phenotype for genetic association tests.

### 2.3 SNP filtering and gene grouping

Genome-wide genotype data were collected at 620,901 markers on the Human610-Quad BeadChip (Illumina, Inc., San Diego, CA). For details on how genetic data were processed, please see Saykin et al., 2010, and Stein et al., 2010a. Different types of markers were genotyped (including copy number probes), but only SNPs were used in this analysis. Several SNPs were excluded from the analysis based on standard filtering criteria, measures used in many other GWAS studies (Wellcome Trust Case Control Consortium, 2007): call rate  $< 95\%$  (42,670 SNPs removed), significant deviation from Hardy-Weinberg equilibrium  $P < 5.7 \times 10^{-7}$  (871 markers removed), autosomal chromosomes only (10,686 SNPs removed), and an Illumina GenCall quality control score of  $< 0.15$  to eliminate “no call” genotypes (variable number of missing genotypes across subjects). We chose to remove SNPs with a minor allele frequency  $< 0.10$  (161,354 SNPs removed) based on our sample size. With our sample of 731 images, we are underpowered to detect associations with SNPs where the minor allele frequency is less than 10%, unless effect sizes are large (Wang et al., 2005; Flint et al., 2010). In addition, excluding SNPs with low minor allele frequencies avoids the risk of finding significant associations where only a small subset of subjects have the rare allele type and do not represent an accurate sampling of the phenotype of interest. If a very low minor allele frequency cutoff is used (e.g.,  $MAF < 0.01$ ) in samples of fewer than a thousand subjects, this may result in cases where an association is driven by a single subject. Clearly, such a result may be unreliable and is unlikely to replicate, so the higher MAF cut-off guards against this.

Due to the filtering based on Illumina GenCall quality control measures, individual subjects have some residual missing genotypes at random SNPs throughout the dataset. Because PCReg requires data without missing genotypes and to maximize the number of subjects included in the analysis, we performed imputation using the software, Mach (version 1.0), to



infer the haplotype phase and automatically impute the missing genotype data (Li et al., 2009). After all rounds of quality control and preparation, 437,607 SNPs remained.

Using the retrieval interface of the software package PLINK (version 1.05; <http://pngu.mgh.harvard.edu/~purcell/plink/>), SNP annotations were made by continuously soliciting the TAMAL database (Hemminger et al., 2006) based chiefly on UCSC genome browser files (Hinrichs et al., 2006), HapMap (Altshuler et al., 2005), and dbSNP (Wheeler et al., 2006). The newly annotated SNPs were grouped by gene, where “gene” is defined by the gene transcript region including both introns and exons. We chose not to include SNPs upstream/downstream from the gene region. This may miss SNPs in promoter or regulatory regions for a gene, but avoids choosing an arbitrary window that may select regulatory SNPs for some genes, but not for other genes whose regulatory regions lie beyond the window length. SNPs that were not located in a gene were excluded (224,057 SNPs removed). All splice variants were considered as belonging to the same gene. After applying SNP filtering criteria, SNP annotation, and gene grouping, 18,044 genes were left for analysis out of the estimated 20,000–25,000 protein coding genes in the human genome (International Human Genome Sequencing Consortium, 2004).

## 2.4 Gene-based association statistics

Independent tests of statistical association with imaging measures were performed for each gene at 31,622 voxels within a whole-brain mask of the MDT across 731 subjects. To test the joint effect of *all* SNPs in a gene on the volume difference at each voxel, we employed a multiple partial-*F* test. A multiple partial-*F* test works by first estimating the fit of a “reduced model” of any number of nuisance variables on a given dependent variable and then estimating the fit of a second “full model” with the nuisance variables and any number of independent variables on the same dependent variable. Each association test results in an *F* statistic, which represents the joint effect of the independent variables on the dependent variable, controlling for the effects of nuisance variables already in the model. The multiple partial-*F* statistic was calculated for each gene at each voxel using equation 1 below. *k* is *df*(full)–*df*(reduced) and *RSS* is the residual sum of squares:

$$F_{k,df(full)} = \frac{RSS(reduced) - RSS(full)}{df(reduced) - df(full)} \bigg/ \frac{RSS(full)}{df(full)} \quad (1)$$

Multiple partial-*F* tests are well suited for testing the effects of multiple predictors on a given phenotype, but genetic data sometimes complicate testing because SNPs in the same gene are often correlated due to high LD. When the SNP values in a cohort of subjects are treated as a vector (whose components are the SNP value in each subject coded in an additive manner: 0, 1, or 2), then the adjacent SNPs can make different subjects’ vectors collinear. The dependence among these almost collinear SNP vectors in the multiple partial-*F* test model can lead to improper signs of beta coefficient estimates, wildly inaccurate magnitudes of beta coefficients, large standard error estimates (Kleinbaum et al., 2008), and false inferences. The reason this occurs is that standard regression models require the inversion of a set of “normal” equations, and when predictors (here SNP vectors) are highly correlated, the equations are not of full rank. This leads to unstable or unreliable solutions. One way out of this predicament is to use a type of regularized or penalized regression, such as ridge regression (also known as sparse regression or Tikhonov regularization), which can be used when there are high correlations among the predictors. Alternatively, dimension reduction may be performed (which we do here), to create a set of predictors that explain the variance in the data but that are no longer correlated.

To avoid the complications of collinearity in the statistical model, we first performed principal component analysis (PCA) on the SNPs within each gene, storing all of the orthonormal basis vectors of the SNP matrix that explained the first 95% of the variance in the set of SNPs. Basis vectors with the highest eigenvalues (higher proportions of explained variance) were included until 95% of the SNP variance was explained, and the rest were discarded. These new “eigenSNPs” approximate the information in the observed SNPs, but lack the collinearity that disrupts the multiple partial-*F* test models. By first performing PCA followed by a multiple partial-*F* test, our method may be considered a variant of PCReg and produces *F* statistics equivalent to those proposed in Wang and Abbott (2008). In this study, the independent variables built into the multiple partial-*F* test full model were the column vector output from PCA performed on each gene with age and sex as covariates. In this way, we tested the joint predictive effect of variation throughout a gene on brain volume variations on a voxel-by-voxel level.

The total number of tests of association for vGeneWAS is very high (18,044 genes  $\times$  31,662 voxels). Because of the massive processing requirement, we coded the PCA and multiple partial-*F* test steps of PCReg using the R statistical package (version 2.9.2; <http://cran.r-project.org/>) using the doMC “multi-core” package (version 1.2.1; <http://www.revolutionanalytics.com/>) to split processing over multiple cores in a single CPU. Processing was parallelized over a cluster of 10 high performance 8-core CPU nodes using the Laboratory of Neuro Imaging (LONI) Pipeline (<http://pipeline.loni.ucla.edu/>). For further data reduction, we only saved data on the gene with the lowest *P*-value at each voxel. This is comparable to our prior work using voxelwise testing of all 500,000 genotyped SNPs, where only the SNP with the lowest *P*-value was retained at each voxel (Stein et al., 2010a). The total time required to complete an analysis was approximately 13 days.

## 2.5 Comparison of SNP-based and Gene-based methods

To examine the situations where PCReg exhibits better (or worse) performance than traditional simple linear regression, we compared the two methods directly on real genetic data. Performing tests on real genetic data as opposed to simulated data is important because the power of each method depends upon the underlying LD structure. Generating simulated data that mimics a chosen LD structure can be just as biased as selecting actual genes, though a significant treatment of the issue of power in PCReg is discussed in Wang and Abbott (2008). We used temporal lobe volume (TLV) summary measures obtained by Stein et al. (2010b), as the phenotype for testing associations for both methods. We performed a genome-wide scan of every SNP from our filtered and annotated genotype data (only including SNPs located within genes) using simple linear regression with SNPs coded following an additive model. We took the top SNP from the analysis, and the rest of the SNPs from the same gene, and performed PCReg on all of the SNPs in that gene. In addition, we performed a gene-wide scan of all of the genes in our dataset using PCReg with SNPs coded following the additive model. We selected the top gene from the analysis and then ran individual tests of association using simple linear regression at each SNP within the top gene. In this way, we were able to compare the performance of each method in cases where the underlying genetic structure might favor one method over the other.

## 2.6 Statistical thresholds and correction for multiple comparisons

As we noted in Stein et al. (2010a), the minimum *P*-value at each voxel, in the null case with *n* independent tests, approximately follows a probability density function (PDF) such that (Ewens and Grant, 2001):

$$f_{\min}(x) = n(1 - x)^{n-1} \quad (2)$$

The PDF derived from equation 2 is a Beta distribution with parameters  $\alpha=1$  and  $\beta=n$ . At each voxel, selecting the minimum  $P$ -value for the top gene then follows a Beta(1,  $n$ ) distribution, where  $n$  is the independent number of genes tested.

However, it is well known that the adjacent SNP values within genes are not statistically independent (Frazer et al., 2007). Genetic loci are inherited in contiguous segments, and some genes co-segregate in blocks. The allele frequencies and structure of genes that co-segregate are more similar than would be expected by chance if they were assumed to be independent. Because of this, the effective number of independent tests ( $M_{\text{eff}}$ ) is less than the total number of tests performed ( $M$ ). By determining  $M_{\text{eff}}$ , we obtain a more accurate estimate of the total number of independent tests performed with vGeneWAS, given the LD structure of our genotype data.

In our sample, we estimated  $M_{\text{eff}}$  by performing permutation tests at three randomly selected, uncorrelated voxels in the brain. We regressed each of the 18,044 genes on the permuted residuals of the reduced model after including the age and sex covariates at each run, and stored the minimum  $P$ -value. Note that, in this case, the phenotype data is null. However, because it is computed from the real data after adjusting for age and sex, the phenotype data (image values) have the same range and statistical distribution as the data tested for genetic associations. By using the genes, one at a time, as regressors on this null data, one can develop a distribution of the resulting  $P$ -values, under null conditions, that can be used to calibrate the significance values that are ascribed to the observed data. As only the minimum  $P$ -value is retained (for the best fitting gene), one can build up a reference distribution for the minimum  $P$ -values, to help gauge the level of surprise in seeing associations in true data. We repeated this process 5000 times at each of the three randomly chosen voxels and merged the data. The distributions of null minimum  $P$ -values from each voxel were nearly identical (Figure 1). Storing the minimum  $P$ -values of the permutation tests yields the expected null Beta distribution given our data. We used a maximum-likelihood function (betafit; Matlab, The Math Works, Inc.) that estimates the best fit for the null Beta distribution by varying the  $\beta$  parameter of Beta(1, $\beta$ ). The value of  $\beta$  approximates the effective number of independent tests ( $M_{\text{eff}}$ ) performed on our data. We then apply a beta inverse transform using the approximated  $\beta$  parameter so that the distribution of  $P$ -values is now on uniform distribution that deviates from the null when there is a signal.

After correcting for the effective number of independent gene-based tests performed at each voxel, we still need to correct for the multiple comparisons across voxels. We used the original false discovery rate method (FDR; Benjamini and Hochberg, 1995), which identifies whether there is any statistical thresholding of the uncorrected  $P$ -value maps that keeps the rate of false positive results within a predefined threshold (we chose 5%, which is conventionally used). This means that, if the results pass the FDR test, approximately 95% of the voxels declared as significant associations will be true positives (averaged over many experiments). In addition, we tested a less conservative variant of FDR, the positive FDR (pFDR), which operates under the condition that at least one true positive finding exists in the data (i.e., one of the null hypotheses is rejected) and yields  $q$ -value correction thresholds similar to the original FDR method (Storey et al., 2003). The pFDR test is implemented in the R statistical package called “qvalue” (Version 1.22.0).

## 2.7 Estimation of expected values in simulated maps

A certain amount of spatial smoothness is expected among voxels in an image. This is most likely explained by the non-independence of volume difference measures at adjacent voxels. Relative volume maps were generated using tensor-based morphometry (TBM), which relies on non-linear registration of each subject's imaging data to a common template. The degree of spatial smoothness in the Jacobian maps derived from the gradient of a deformation field



depends on the choice of the regularizer used by the warping algorithm (Laplacian, elastic, fluid, sKL, etc.) and on the resolution of numerical grid chosen to solve the differential equations whose solution is the deformation field. Volume difference maps based on the deformation field vectors are spatially smooth, as are any resulting statistical maps. In addition to image smoothness, certain noncontiguous voxels in distant regions of the brain can have surprising covariance patterns despite their spatial separation (Fillard et al., 2006).

We examined whether the size of voxel clusters associated with the same gene from our vGeneWAS analysis differed from the cluster sizes expected under the null hypothesis of no association at all, given the non-independence of signals at adjacent voxels in our images. In addition, we wanted to determine whether the number of unique, top genes from across the brain significantly differed from the number of top genes expected by chance. We generated simulated cluster maps by first creating a correlation matrix of  $r$  values representing the Pearson's correlation between any given voxel and all other voxels in the brain. Next, we randomly selected (without replacement) a voxel,  $V_s$ , and all corresponding voxels with an  $r^2$  value (proportion of variance explained) greater than 0.8 from the correlation matrix. We chose to only select voxels from the correlation matrix with an  $r^2$  greater than 0.8 because this provided the largest cluster size estimates in the simulated output maps. The  $r^2$  value of each voxel-to-voxel relationship was then used to divide the interval [0,1] of a uniform distribution such that the correlation between a voxel and  $V_s$  was directly related to the area under the curve occupied by that voxel's area under the curve or "bucket." The size of each voxel's bucket was recalculated each time we chose a new  $V_s$ . We selected a random number on the interval [0,1], and, depending on its value (which "bucket" it fell in), assigned the same categorical variable link (e.g., a, b, c) to  $V_s$  and the voxel whose bucket was selected from the uniform distribution. This linked the two voxels. The probability of a given voxel being chosen from the uniform distribution was directly related to how correlated it was to  $V_s$ . We continued the process by randomly choosing a new  $V_s$  from the correlation matrix. If a randomly selected voxel  $V_s$  did not contain a linking variable, but selected a voxel from the uniform distribution that already contained a link, then  $V_s$  was assigned the linking variable of the voxel selected from the uniform distribution. If  $V_s$  already contained a linking variable and the voxel chosen from the uniform distribution had not previously been assigned a variable, the two voxels were linked using the existing linking variable of  $V_s$ . If both  $V_s$  and the voxel selected from the uniform distribution already contained a linking variable, we kept each variable as-is and then continued with the process. Finally, voxels that were not correlated to any other voxels in the image, with an  $r^2$  value greater than 0.8, were assigned a non-linking random variable. After iterating through every voxel in the image, each voxel had a categorical variable that either linked it to other voxels or only to itself. We ran this entire simulation process 100 times, generating a new simulated cluster map each time. By considering the correlation of a given voxel to all other voxels in the image, as opposed to using a single summary measure of smoothness throughout an image, we were able to model the expected clustering among adjacent voxels and non-independent, spatially separated clusters.

Based on the 3D pattern of voxels and the variables linking voxels together, we used a nearest neighbor algorithm to measure cluster sizes of adjacent voxels with the same linking variable value. Using the cluster size estimates from each simulated map, we were able to determine the expected distribution of cluster sizes based directly on our study dataset. In addition, we used the total number of unique linking variables in each simulation as an estimate of the number of independent voxels in our dataset. Because non-independent, correlated voxels may tend to be associated with the same gene, we can use the total number of independent voxels to estimate the number of top associated genes we would expect to find in null cluster maps made from our actual test data. We used the estimated number of independent voxels,  $V_i$ , to randomly select (with replacement) a gene from the set of 18,044

genes and repeated the selection  $V_i$  times. We found the number of unique genes represented for each simulated output map and then took the average.

### 3. Results

#### 3.1 Comparison of methods

To examine the differences between gene-based and SNP-based association methods (which are more standard), we compared the results of PCReg to linear regression using temporal lobe volume (TLV) data from a previous study (Stein et al., 2010b) as the phenotype. We chose to focus on the top gene or SNP identified by each method in order to examine performance when the variant chosen is deliberately selected to favor one of the two methods. *GRIN2B* was identified as the gene with the SNP variant that was most significantly associated with TLV ( $P = 4.03 \times 10^{-7}$ ). We identified each of the 129 SNPs within the *GRIN2B* gene, and then performed linear regression at each SNP and PCReg as a single gene test with TLV as the phenotype. The  $-\log_{10}(P\text{-value})$  of each SNP-based test is shown in Figure 2a with single gene test results overlaid (black dotted line). It is clear that the main effect detected with linear regression is much greater in this case. It is important to note that we tested each of the 129 SNPs within the *GRIN2B* gene, which would require any significant  $P$ -values identified to be corrected for multiple comparisons before further study. In this example, however, there are several SNPs that beat the Bonferroni-corrected significance threshold ( $\alpha = 3.8 \times 10^{-3}$ ). In comparison, the gene-based test of *GRIN2B* using PCReg was a single test not requiring correction for multiple comparisons and maintained a nominal significance value ( $P = 0.012$ ). Also, we compared *BEST3* - the gene identified to be most significantly associated with TLV via PCReg - with the linear regression output of each SNP within the gene (Figure 2b). The significance of the main effect of the gene-based test is much stronger ( $P = 2.9 \times 10^{-4}$ ) than the best linear regression result ( $P = 0.063$ ). This demonstrates a case where variance components from individual markers that are not significant via linear regression may be combined into a single significant test statistic.

#### 3.2 Relation of gene significance to number of SNPs

Wang and Abbott examined whether the power to detect associations in genetic data is influenced by the number of eigenSNPs included in a PCReg model. They found that models with greater numbers of eigenSNPs do not have increased power to detect associations (Wang and Abbott, 2008). However, each additional eigenSNP included in the model uses a degree of freedom. It is then possible that PCReg and similar regression methods are biased toward selecting effects of smaller versus larger genes (Chapman and Whittaker, 2009). We examined our results for gene-size bias and verified that the number of eigenSNPs in the PCReg model within the top genes from our run of vGeneWAS was not correlated with the observed  $P$ -value, using a Pearson's product-moment correlation test ( $r = 0.0045$ ;  $P = 0.42$ ). In addition, we verified that the number of eigenSNPs in each of the 18,044 genes at a single voxel was not correlated with its significance level ( $r = 0.0066$ ;  $P = 0.29$ ). We also compared the number of eigenSNPs in each gene (mean and median: 14.3 and 9) with the number of eigenSNPs in the top genes from our analysis (mean and median: 13.6 and 5). It remains possible that we missed effects of very large genes, but this is inevitable in small samples as the number of eigenSNPs needed to adequately encode the majority of variation in large genes tends to approach the sample sizes, reducing the available numbers of degrees of freedom for the whole-gene tests.

#### 3.3 Voxelwise GeneWAS

We generated maps of significance where each color-coded voxel in the brain shows the  $P$ -value of the most highly associated gene at that voxel (Figure 3). There are several spatially contiguous regions throughout the brain with raw minimum  $P$ -values lower than  $P < 10^{-7}$ . In

addition, some of the top genes identified show symmetric clustering across hemispheres. Brain structures are highly symmetric between hemispheres, at least for most brain regions, so symmetric genetic associations may be biologically plausible because the volumes of symmetric structures co-vary across subjects, so they may share similar genetic determinants. However, evidence of symmetric patterns of association in the brain does not necessarily imply biological plausibility (Fillard et al., 2006).

We used a simulation-based test to build the expected null distribution of cluster sizes given our image data. We compared the distribution of the cluster size values in simulated maps to the cluster sizes obtained from vGeneWAS. The proportion of the null (simulated) maps that contained small clusters is much greater than in vGeneWAS, while the proportion of the vGeneWAS map that contained large clusters was greater than in the null maps (Figure 4). The minimum and maximum cluster sizes for the simulated maps were 1 and 14 voxels (64 and 896 mm<sup>3</sup>), respectively. The minimum and maximum cluster sizes for vGeneWAS were 1 and 429 voxels (64 and 27456 mm<sup>3</sup>). This demonstrates that a large proportion of clusters of voxels associated with the same top gene are larger than would be expected based on completely null data, even taking into account the non-independence of voxels in our dataset.

Based on our simulated cluster maps, we used the number of unique clusters as an estimate of the number of independent voxels. The estimate of the number of independent voxels based on 100 runs of the simulation tests was  $11900.8 \pm 50.6$  (mean  $\pm$  standard deviation) out of the 31662 total voxels. We performed tests to estimate the number of genes we should expect to find in our analysis based on the non-independence of voxels in our data. We used the number of independent voxels estimated from the simulation to randomly select (with replacement) from our list of 18,044 genes. We tallied the number of unique genes represented for each simulated cluster map and found the average was  $8721.4 \pm 44.9$  (mean  $\pm$  standard deviation). We measured the total number of unique genes to be 5333 from our run of voxelwise GeneWAS. The number of observed genes is significantly lower than the number of genes expected based on the null cluster maps ( $P < 0.01$ ). Combined with our cluster size comparisons, this suggests that the top genes identified in our analysis tend to have a much more broadly distributed effect than expected based on null data, even taking into account the intrinsic spatial non-independence of our data. The top 20 genes most significantly associated with any voxel are listed in Table 1.

The GRB-associated binding protein 2 gene, *GAB2*, is the most significantly associated gene in our analysis and has previously been linked to late-onset Alzheimer's disease (LOAD) (Reiman et al., 2007). Reiman et al. (2007) identified 10 SNPs from the *GAB2* gene that were significantly associated with LOAD and *APOE* allele status in 1411 cases and controls from 20 NIA-sponsored Alzheimer's Disease Centers. Replication attempts in independent samples have yielded mixed results (Ramirez-Lorca et al., 2009; Lin et al., 2010; Chapuis et al., 2008), but large meta-analyses of several databases shows that *GAB2* may indeed have a moderate effect on the development of LOAD (Ikram et al., 2009; Schjeide et al., 2009). Specifically, the meta-analysis of the marker rs2373115 in nine studies has an odds ratio of 0.85 and a 95% confidence interval for the odds ratio of [0.76, 0.94]. In addition, the AlzGene website lists *GAB2* as being in the top 20 genes likely related to Alzheimer's disease (September 3, 2010; <http://www.alzgene.org/>). *In vivo* testing shows that *GAB2* is over-expressed in certain brain regions such as the hippocampus and posterior cingulate cortex in patients with LOAD (Reiman et al., 2007). Experiments with small-interfering RNA (siRNA) and *GAB2* reveal that the normal function of *GAB2* proteins prevents the formation of serine-262 phosphorylated tau tangles (Reiman et al., 2007). No studies, to our knowledge, have considered morphometric effects of *GAB2*

variants. The *GAB2* associations show a symmetric signal in the white matter superior to the lateral ventricles (Figure 5).

The second most highly associated gene, leucine-rich repeat and death domain containing protein (*LRDD*), is expressed in the brain and may mediate cell apoptosis and DNA repair (Telliez et al., 2000). In addition, *LRDD* has been implicated in the p53 tumor-suppression pathway likely by signaling cell apoptosis in response to DNA damage (Brown et al., 2009). *LRDD* was the most significantly associated gene in a cluster of voxels in a white matter tract of the occipital lobe, possibly the optic radiations (Figure 5).

Associations with protein tyrosine phosphatase receptor type beta, *PTPRB*, are detected in the cerebellum (Figure 5). *PTPRB* interacts with neural receptors and cell-adhesion molecules and is involved in neurite development and neuronal differentiation (Ishiguro et al., 2008). *PTPRB* has also previously been associated with alcohol and drug abuse via genome-wide search (Ishiguro et al., 2008). In addition, an expression study found that *PTPRB* encoded proteins are present in the gastric mucus and other tissues of gastric cancer patients (Wu et al., 2006).

The fourth and fifth most significantly associated genes are zinc finger protein 462 (*ZNF462*) and immunoglobulin superfamily member 5 (*IGSF5*), respectively. *ZNF462* is the most significantly associated gene in a cluster of voxels in the upper-left grey matter of the parietal lobe (Figure 5). Interestingly, *IGSF5* shows symmetrical clusters of association in the temporal lobe and the surrounding cerebrospinal fluid (CSF) at the base of the brain (Figure 5). Neither gene is well studied, but *IGSF5* may be involved with junction cell adhesion (Hirabayashi et al., 2003).

Other genes of interest identified in our analysis include *ARALAR*, which encodes a calcium-binding mitochondrial protein that is highly expressed in the brain (del Arco and Satrustegui, 1998). *ARALAR* has previously been associated with autism (Ramos et al., 2004), but the claims are controversial (Rabionet et al., 2006). *CHRM5*, is a muscarinic acetylcholine receptor M5 coding gene and has previously been associated with schizophrenia (De Luca et al., 2004). *S100B*, encodes a zinc-binding protein over-expressed in patients with Alzheimer's disease and interacts with Tau proteins (Yu et al., 2001).

### 3.4 Correction for multiple comparisons

After permutation testing to determine the effective number of independent gene tests, we need to model the function parameters so that we can transform the data for correction for multiple comparisons. The effective number of independent tests was estimated to be 15,636, which is a moderate reduction from the 18,044 genes measured directly in this experiment. We therefore chose to model the null distribution as Beta(1, 15636). The probability density function (PDF) of Beta(1, 15636) on the normalized histogram of observed *P*-values fits the data well with only small deviations from the original Beta(1, 18044) (Figure 6a). We note, however, that our estimate is determined both by SNP density and degree of coverage in the SNP marking scheme of our study. Experiments that use different SNP-chips, include sex chromosomes, or use different annotation methods may encounter different estimates. To determine how well the expected null distribution compares to the observed PDF, we compare each distribution directly in a Q-Q plot (Figure 6b). The expected null distribution also fits the observed data well.

*P*-values suitable for multiple comparison correction via FDR methods should have a probability distribution on the interval [0,1] that is uniform in the null case, i.e., its cumulative distribution is diagonal in the null case (Benjamini and Hochberg, 1995; this is the basis for the false discovery rate method). Our Beta-distributed experimental *P*-values

need to be corrected so that they meet the assumptions of the FDR model. Using the analytic  $\beta$  parameter from the null Beta distribution, we fit a cumulative distribution function (CDF) to our observed data yielding a new distribution of corrected  $P$ -values that deviate from the uniform distribution only when the data are not null. A histogram of the observed corrected  $P$ -values (Figure 6c) shows that the cumulative distribution is approximately equivalent to a uniform distribution. A Q-Q plot of the expected null distribution corrected  $P$ -values against the observed corrected  $P$ -values shows that the two distributions differ (Figure 6d). A Q-Q plot of two identical distributions will lie on the null 45-degree diagonal line ( $y = x$ ). There are two things that can cause a Q-Q plot to deviate from the null: incorrectly modeling a distribution or significant data. The line representing the observed compared to the expected results in the Q-Q plot in Figure 6d is steeper than and deviates from the null line at lower  $P$ -values. This suggests that the distribution of  $P_c$  values is left skewed and more dispersed than the theoretical uniform distribution (Thode et al., 2002). It is possible to apply a further correction to our observed  $P_c$  distribution by using a Q-Q plot of an analytic null distribution versus the theoretical uniform distribution as a hash table. However, we are only selecting the genes with the lowest  $P$ -value at each voxel so monotonic  $P$ -value corrections will not change the distribution of  $P_c$ -values.

We used two methods to control the FDR of the corrected  $P$ -values ( $P_c$ ). We used the original FDR method (Benjamini and Hochberg, 1995), which appropriately controls for multiple comparisons when the covariance of test statistics shows a positive regression dependency (Benjamini and Yekutieli, 2001). We found that the false discovery rate for the second most highly associated gene in our results (*LRDD*) could only be controlled at a threshold of  $q = 0.30$  (i.e., allowing a 30% false discovery rate) after applying a statistical threshold of  $P_c = 5.36 \times 10^{-4}$ . In addition, the pFDR  $q$ -value threshold (Storey et al., 2003) was  $q = 0.23$  for the most significantly associated gene at any voxel (*GAB2*). In other words, the vGeneWAS results could not be controlled at the conventional false discovery rate, but show promise.

### 3.5 Post hoc analysis

Voxelwise GeneWAS results in a map that shows only the top gene at each voxel. The top gene at each voxel may be the most significant gene in a certain region, but it may also have a more distributed effect throughout the brain, with effects in additional regions where it is not the top gene. In addition, genes that do not have a large main effect might never be selected in this type of analysis, but still could have a large distributed effect on the brain.

We tested the effect of the top gene in our analysis, *GAB2*, at every voxel across the brain using PCReg. We stored each  $P$ -value in a map and applied the original FDR method. Voxels surviving the FDR threshold are shown in Figure 7. These are *post hoc* tests, so are exploratory, and require replication in independent samples, but it is quite clear that *GAB2* has a much greater distributed effect on the brain than could be determined from the vGeneWAS results. Future implementations of vGeneWAS might consider the effects of multiple genes at a voxel to account for the case where a gene is significant in its effect of explaining variations in the image, but is not necessarily the top gene. In addition, vGeneWAS could be further improved by considering the distributed effects of genes. If a gene has an effect over a large region, but is not the top voxel, it will be completely overlooked in the current implementation of vGeneWAS. Adaptation of cluster-level inference to these maps would be of interest, as well as tests that combine cluster extent and height (Hayasaka and Nichols, 2004). Existing adaptations of the original FDR method, such as “searchlight” FDR, could be useful here as it produces region correction thresholds that are sensitive to small clusters of positive signals in imaging data, but appropriately conservative in its correction of false positives (Langers et al., 2007).



To understand the contributions of each individual SNP in the *GAB2* gene, we performed *post hoc* association tests for each SNP with the phenotype value from the top voxel in the brain. It should be noted, however, that choosing the *GAB2* gene to compare the results of SNP-based linear regression with gene-based PCReg provides *P*-values which are biased by the previous gene-wide brain-wide search because *GAB2* was identified using PCReg. There were 20 SNPs from the *GAB2* gene in our genotyped data. Of these, only three passed the nominal significance level in SNP-based association tests ( $P = 0.05$ ). The most significant SNP identified, rs7927923, has  $P = 9.1 \times 10^{-3}$ . The other two significant SNPs, rs1981405 and rs1893447, showed effects with  $P = 0.027$  and  $P = 0.049$ , respectively. A total of 16 SNPs out of 20 are in high LD with the most significantly associated SNP ( $r^2 > 0.6$ ). Only one of the SNPs from our analysis overlapped with SNPs used in previous *GAB2* association studies, most likely because we are using different genotyping platforms. Clearly, the gene-based test was more powerful at detecting an association in this case than each SNP tested individually (compare the dotted line with the colored dots in Figure 8).

### 3.6 Power comparisons

To assess the differences in power afforded by vGeneWAS relative to existing SNP-based methods, we compared the  $P_c$ -values from vGWAS obtained in our previous study by Stein et al. (2010a), with the  $P_c$ -values resulting from vGeneWAS (Figure 9). The proportion of  $P_c$ -values greater than a given FDR threshold for each method is directly related to differences in effect sizes. The FDR of the results from vGWAS could only be controlled at a threshold value of  $q = 0.50$ , whereas the FDR threshold for vGeneWAS is somewhat lower, although not passing the conventional FDR level ( $q = 0.30$ ; Figure 9). This suggests that the vGeneWAS method may have more power, in principle, to detect genetic associations, although neither test controlled the false discovery rate at the conventional level.

## 4. Discussion

### 4.1 Methodological overview

Here we present a method to conduct a voxelwise gene-wide association study (vGeneWAS), testing the aggregate effect of multiple SNPs within each gene. In summary, (1) we implemented a gene-based association test using principal components regression (PCReg); (2) we performed association tests at every voxel within a full brain mask where the value at each voxel was the local volume difference relative to the mean template while controlling for age and sex; (3) we generated a beta distribution of *P*-values by selecting only the most significant gene at each voxel; (4) with permutation tests, we estimated the effective number of tests performed; (5) we corrected for multiple comparisons in a two step procedure - we estimated  $\beta$  using the CDF of the analytic Beta distribution and then corrected the new uniform distribution using two different FDR methods.

None of the genes identified passed the standard FDR threshold ( $q = 0.05$ ). However, many of the genes identified have previously been associated with brain differences or disorders. The top gene identified is a known Alzheimer's risk gene, *GAB2*, lending plausibility to the method. Many of the genes identified are highly expressed in the brain or differentially expressed, depending on disease status. The findings in this study warrant further examination and replication attempts.

### 4.2 Assessment of the model

Our method selects the top gene at each voxel, to reduce the amount of data. Choosing only the top gene at each voxel, however, can hinder the extensibility of our results. This may miss many genes with distributed effects, if the main effect of the gene is never the largest at

any voxel. Future implementations of vGeneWAS could consider the relationship among voxels when performing association tests. Liu et al. (2009) used parallel ICA to relate brain network data from fMRI to SNP data. They selected only a small set of 367 predefined SNPs based on a set of candidate genes for schizophrenia; this does not leverage all of the available data in the genome. Similar approaches have been attempted on voxel-based morphometry (VBM) data from structural MRI (Jagannathan et al., 2010). However, this approach used the same subset of SNPs used by Liu et al. (2009). Vounou et al. (2010) proposed a method called sparse reduced-rank regression (sRRR) which uses the sparseness of signals to simultaneously select phenotypes and genotypes. Power estimates suggest that sRRR is more powerful than using individual tests at each voxel; this may prove to be very useful in the future.

Principal components regression (PCReg) is an efficient method to test the joint association of all SNPs within a gene simultaneously. PCReg detected associations with genes missed by SNP-based regression (Figure 2b). By leveraging the LD in a gene, PCReg encodes variance throughout a gene to test for associations. We identified situations where SNP-based regression models may have more power (Figure 2a). If a single SNP has a large main effect, then testing the joint effect of all SNPs within that gene may dilute the association; the cumulative *P*-value from gene-based tests may be lower. However, when one considers the drastic reduction in the number of independent tests when comparing SNP-based linear regression with PCReg, gene-based testing offers advantages.

Another concern with PCReg and related regression models is that each predictor added to the regression model consumes a degree of freedom. There may then be some detection bias in the regression model, where smaller genes are found to be more significantly associated with the phenotype than larger genes, because the regression models of small genes have more degrees of freedom (Chapman and Whittaker, 2009). While we did not observe this effect, it is an important factor to consider when interpreting results. Additionally, SNPs combined into a single test statistic in PCReg could have different directions of effects, disrupting the power to detect an association. However, the situation where a gene contains SNPs with negative correlations with respect to the phenotype may be relatively rare as it requires two nearby loci to be in LD with different causal variants (Chapman and Whittaker, 2009).

Other multivariate regression methods may offer greater power to detect genetic associations than PCReg, which is used here as an example. Wang and Elston (2007) compress genome-wide genotyping data across subjects using a Fourier transform, and assign weights to the low-frequency components in a regression model. This method is similar to PCReg, as it collapses the number of genetic tests performed while capturing much of the variance across markers. Kernel-based methods have been implemented as non-parametric gene-based tests to increase power over SNP-based methods; however, these methods have only been implemented in case-control studies (Mukhopadhyay et al., 2010). Ridge regression (Malo et al., 2008; Sun et al., 2009) and lasso-based penalized regression (Zou et al., 2006; Wu et al., 2009) can both powerfully detect associations in genetic data. In fact, direct comparisons between ridge regression, lasso-based penalized regression, and PCReg show that the first two methods may be more powerful than PCReg depending on the underlying genetic architecture (Bovelstad et al., 2007; Benner et al., 2010). However, ridge regression and especially lasso-based penalized regression are extremely computationally intensive compared to PCReg. There is a huge computational requirement to complete a vGeneWAS analysis, which searches the whole image in addition to the whole genome. Due to this, we decided to strike a balance between power and computational complexity to complete the analysis in a feasible time frame. Future implementations of vGeneWAS could be improved

by using additional multivariate regression methods, although they may need to be modified for speed.

A current limitation of our method, as described here, is that in its current form family-based designs (such as pedigree structures) cannot be used. The patterns in allele frequencies in a family cohort depend on kinship, and mixed-effects models would be required to control for kinship structure. Several such methods exist for SNP-based linear regression, such as Efficient Mixed-Model Association (EMMA; Kang et al., 2008). However, to the best of our knowledge, multivariate mixed-model regression has not been attempted in a genetic context. One multivariate gene-based method called versatile gene-based association (VEGAS) avoids multivariate mixed-model regression by converting SNP-based  $P$ -values into a multivariate gene-based test statistic (Liu et al., 2010). As there are already many methods for SNP-based mixed-model regression, VEGAS is aptly suited to perform gene-based association tests in family-based populations. As the VEGAS test statistic is determined by SNP-based  $P$ -values, it will not be able to detect associations where the cumulative  $P$ -values are not significant. In this way, if a series of SNPs contribute a small amount of variance in a gene, VEGAS will miss them, because the SNP-based method will as well.

While this is one of the largest imaging genetics studies to date, our sample size still may be underpowered to detect moderate effects of genetic variants on the brain. Future studies in imaging genetics are likely to benefit from meta-analysis, which aggregates GWAS results from multiple cohorts to determine reproducible genetic associations. This aggregation of large datasets can be used to boost power to detect SNPs with smaller effects. One such effort now underway is the ENIGMA project (ENIGMA Consortium, 2011). In cases where there is not enough data available to perform a true meta-analysis, discovery and replication datasets may be useful. In early tests using brain images, some genetic associations seen in a discovery cohort have been replicated in independent samples (e.g., Rajagopalan et al., 2011). However, our main purpose in this study is to demonstrate a method to conduct voxelwise, gene-based analyses, which will become more powerful as imaging databases continue to grow rapidly worldwide in size and content. In assessing whether our results may generalize to new datasets, we note that we examined only a tight age range in our study (elderly subjects), and this may affect the genes found to have morphometric effects on the brain. If the genes have a varying expression pattern over time some of the top genes detected in our analysis may not be dominant during other parts of the lifespan. Although we controlled for age effects on brain structure in our analysis, it is still unknown whether the identified genetic associations with brain morphology are under some age-related influence; it is also not known if these genes are expressed in a typical/atypical age-related fashion. Datasets drawn from different parts of the lifespan would offer maximal power to detect genetic variants relevant for brain structure (as discussed in Braskie et al., 2011; see Rajagopalan et al., 2011, for an example).

One limitation of GWAS analyses is that they overlook rare variants, which are also emerging as key players in the genetic underpinnings of mental disorders (Bansal et al., 2010). Our method does not consider these rare variants, but may play an important role in explaining variance in complex traits that is not accounted for by common variants. Examination of rare variants is still relatively costly (as it requires deep sequencing of large numbers of subjects). Some types of rare variant can be genotyped on SNP-chip platforms (such as copy number variants) they require separate analytical techniques from those considered in this paper (Bansal et al., 2010). Each of these limitations will be more feasible to address when the cost of deep sequencing drops and sample sizes are large enough to reliably implicate many genes simultaneously.

Our gene-based test may be more powerful than univariate methods in certain cases, but not always. The five top genes identified in the present study do have some biological plausibility; some are known to be expressed in the brain and implicated in brain disorders. However, there are other genes missing in the short list of genes in the present findings that are frequently found using univariate approaches and strongly implicated in complex behavioral pathologies across mammalian species, such as the *BDNF* val66met substitution.

To explain this, we note that the analysis of currently available imaging genetics data is very underpowered. In addition, sample sizes needed to reliably detect a genetic association are even greater when multiple genes are assessed or when genome-wide search is conducted. By contrast, the *BDNF* val66met substitution is often treated as a candidate gene, and if that is done, it is conventionally agreed that its effects must only satisfy a nominal significance level of  $P < 0.05$ , if no other genes or SNPs within it are tested. In our own work on a different cohort scanned with DTI (Chiang et al., 2010a), we were able to detect robust associations between the *BDNF* val66met polymorphism and fiber integrity (fractional anisotropy) assessed with DTI. We were also able to replicate these associations in two non-overlapping (independent) samples of subjects. Even so, the significance level used ( $P < 0.05$ ) was far more lenient than the very strict thresholds required to control for false positives when the whole genome is searched. As such, false negative results in a GWAS study (e.g., not finding a significant association of the *BDNF* val66met substitution when it is in fact a true association) does not mean that a gene does not affect a phenotype, just that an association was not detected at the very stringent statistical cutoff applied to account for multiple comparisons across the genome.

A review of the literature also suggests that *BDNF* val66met, while a popular target of study, has a mixed history. The *BDNF* val66met substitution has been inconsistently implicated in mood disorders, Alzheimer's disease, and quantitative measures of memory (Bagnoli et al., 2004; Combarros et al., 2004; Nacmias et al., 2004; Nishimura et al., 2004; Tsai et al., 2004; Desai et al., 2005; Matsushita et al., 2005; Vepsäläinen et al., 2005). In a secondary, *post hoc*, analysis, we tested the effect of the *BDNF* val66met allele in this current sample using standard univariate regression (with a dominant model, controlling for age and sex) and it did not survive correction for multiple spatial comparisons. An association test of the *BDNF* val66met allele with the hippocampal volume of the 731 subjects in this dataset did not survive the nominal significance level of  $P = 0.05$ . As such, we were not able to use *BDNF* as a “gold standard” gene; arguably, there are not yet any such genes with universally replicable associations on brain structure that can be used to gauge the face validity of novel methods.

Another limitation is that any approach that stops after selecting only one gene per voxel is not biologically plausible as a model of phenotypes with complex genetic determination. As such, the development of gene-based tests should be considered as a way station towards a more sophisticated treatment of complex genetic effects, in pathway or gene-gene interaction models (Inkster et al., 2010). However, testing gene-gene interactions in the vGeneWAS framework is computationally intractable as the number of tests quickly approaches  ${}^nC_k = n!/((n-k)!k!)$  at each voxel in the whole brain to test for interactions among all sets of  $k$  different genes drawn from an overall pool of  $N$  genes. Additionally, interaction effect sizes are generally much smaller than the main effects we are searching for in this paper, so our sample size would need to be much larger to accommodate the necessary correction for multiple comparisons and smaller effect sizes (Cordell, 2009). As genome-wide interaction analysis (GWIA; Marchini et al., 2005) is computationally intensive and underpowered with current imaging datasets, we recently developed an alternative method (Chiang et al., 2011b,c) to detect likely gene-gene interactions among SNPs, without having to compute all  $N(N-1)/2$  pairwise or all  $n!/((n-k)!k!)$   $k$ th-order interactions on the genome.

Two advantages of this genetic network analysis, relative to genome-wide interaction analysis (GWIA) are apparent: (1) genetic correlation can tap into the natural latent structure of gene action in a brain image; (2) voxel clustering by genetic affinity leads to high power to find SNPs with correlated effects in genome-wide scans.

### 4.3 Biological significance of the findings

*Gene*-based tests of association across the genome and brain have not been attempted before, to our knowledge. Recently, imaging genetics studies have focused on single-locus associations with summary brain volume measures or 3D statistical parametric maps. vGeneWAS advances the burgeoning field of imaging genetics by providing the framework to perform multivariate, gene-based association tests. It does not restrict analyses by requiring prior hypotheses about a specific causal variant or *ad hoc* region of interest. vGeneWAS is the first attempt to apply gene-based tests to morphometric imaging data and opens up more possibilities to discover putative genetic variants that contribute to differences in brain structure. This may help when the main effect of each variant in a gene is too small to detect with traditional SNP-based methods.

Although vGeneWAS is a multivariate, gene-based method, we identified genes previously associated with brain disease using SNP-based tests. Many variants in the *GAB2* gene are implicated in the development of late-onset Alzheimer's disease (LOAD) and are thought to interact with the *APOE* epsilon 4 allele. In the pattern of effects for *GAB2* on the brain (Figure 7), the highlighted areas are generally periventricular, and ventricular enlargement is a prominent characteristic of AD (de Leon et al., 1989; Chou et al., 2010). As we noted in our prior papers on TBM in Alzheimer's disease (Hua et al., 2008), there is occasionally a ring of voxels around the lateral ventricles that show partial volume effects that mostly like reflect ventricular expansion. Clearly, the ventricular expansion itself indirectly results from the diffuse loss of brain parenchyma, so the changes detected there may also reflect, to some extent, atrophic processes more remote than the voxels singled out in the voxel-based maps. *LRDD* is highly expressed in the brain and is involved with DNA repair including signaling apoptosis in tumor cells. *PTPRB* is associated with addiction to drugs and alcohol and may be involved with tumor regulation (Telliez et al., 2000; Wu et al., 2006; Ishiguro et al., 2008; Brown et al., 2009). Based on gene expression and links to brain diseases, many of the genes identified in our analysis may have differential morphometric effects across the brain. In addition to some of the more well-studied genes, we identified many genes such as *IGSF5* and *ZNF462* that have little research available to infer their plausibility. However, almost all of the genes identified in our analysis are highly expressed in the brain, which at least suggests that the genes may have a role in brain function. Further analysis is required to examine to what extent each gene variant identified in our analysis mediates brain differences.

Many of the associations identified here seem to have a plausible story, but we need to consider that some of the patterns of association, especially clusters of association, may be due to short-range spatial correlations in the images. Adjacent voxels in brain scans tend to covary, as do Jacobian maps used to represent a localized measure of volume difference. These methods rely on non-linear algorithms that generate spatially smooth deformations. In addition to the simulated (null) cluster size maps in this study, Stein et al. (2010a) found that a small amount of spatial clustering is seen even if the genetic data is null. Also, voxels were down-sampled which may introduce partial volume effects. However, performing a vGeneWAS scan on non-down-sampled, original sized images is estimated to take 4,372 days (or approximately 12 years) to complete. To this end, the extent to which a gene affects regions of the brain should be interpreted cautiously; however, certain patterns of gene effects that appear in non-adjacent structures or in large clusters may signal gene effects not attributable purely to spatial smoothing or partial volume effects.



To better understand the contribution of genes to global versus local brain structure differences, we conducted association analysis on both (1) the globally normalized brain images and (2) estimates of total intracranial volume (eTIV) that contain information on overall differences in brain scale. We searched for specific gene effects on global brain volume differences using our gene-based method. We computed brain volume measures (eTIV) from our dataset using the automated FreeSurfer package (Fischl et al., 2002). Using the eTIV measure as the phenotype we tested each of the 18,044 genes for association. Looking at the top genes that we found in our analysis of the normalized images, none of the top 20 most highly associated genes was associated with eTIV phenotype. This provides further evidence that the genetic effects we are detecting exert influences on regional brain volumes rather than simply reflecting non-specific effects on the overall volume of the brain.

Additionally, results should be interpreted cautiously when global anatomical normalization is used. By removing, as far as possible, the effects of individual brain size variation from the data, it is possible to discover genes that may have a specific effect on a particular structure, above and beyond any overall genetic effects on brain size (as brain size itself is heritable). Global normalization is commonly performed in all brain mapping studies, as many extraneous factors affect an individual's head size, body size or height that may not be relevant for cognition or for understanding brain function. Global anatomical normalization adjusts for this source of variance in the data, to a large extent, making more localized effects easier to identify. Even so, by applying global anatomical normalization, some genes may be missed that influence the total size of the brain. In fact, if a gene were responsible for influencing brain size, but had a uniform effect on all brain regions, it would be missed in the current analysis, as global effects are discounted. As such, in addition to mapping gene effects, it makes sense to also perform genetic analyses of whole brain size, as has been performed in two recent studies (Paus et al., 2011; ENIGMA Consortium, 2011).

In conclusion, our method may be used to perform gene-based tests on any 3D brain maps, such as data from voxel-based morphometry, diffusion tensor imaging, and cortical surface data. In addition, we found a set of candidate genes that may substantially affect brain morphometry and warrant further study.

## Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904, 3U01AG024904-03S5). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. We also thank the many contributors to ADNI-1 genotyping sample curation at NCRAD (Kelley Faber), performing BeadChip assays at TGen (David Craig), and bioinformatics problem solving (Indiana U: Kwangsik Nho; UC Irvine: Anita Lakatos, Guia Guffanti; Pfizer: Bryan DeChairo).

## References

1. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P, Gibbs RA, Belmont JW, Boudreau A, Leal SM, Hardenbol P, Pasternak S, Wheeler DA, Willis TD, Yu FL, Yang HM, Zeng CQ, Gao Y, Hu HR, Hu WT, Li CH, Lin W, Liu SQ, Pan H, Tang XL, Wang J, Wang W, Yu J,

- Zhang B, Zhang QR, Zhao HB, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Shen Y, Yao ZJ, Huang W, Chu X, He YG, Jin L, Liu YF, Shen YY, Sun WW, Wang HF, Wang Y, Wang Y, Wang Y, Xiong XY, Xu L, Wayne MMY, Tsui SKW, Xue H, Wong JTF, Galver ILM, Fan JB, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Kwok PY, Cai DM, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Sham PC, Song YQ, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Fry B, Maller J, McCarroll S, Patterson N, Pe'er I, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Varilly P, Stein LD, Krishnan L, Smith AV, Thorisson GA, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan WH, Munro HM, Qin ZHS, Thomas DJ, McVean G, Bottolo L, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Zhang HC, Zeng CQ, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwdimmah C, Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu RZ, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Richter DJ, Ziaugra L, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, Archeveque PL, Bellemare G, Saeki K, Wang HG, An DC, Fu HB, Li Q, Wang Z, Wang RW, Holden AL, Brooks LD, McEwen JE, Bird CR, Guyer MS, Nailer PJ, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Witonsky J, Zacharia LF, Kennedy K, Jamieson R, Stewart J, Consortium IH. A haplotype map of the human genome. *Nature*. 2005; 437:1299–1320. [PubMed: 16255080]
2. Bagnoli S, Nacmias B, Tedde A, Guarnieri BM, Cellini E, Petrucci C, Bartoli A, Ortenzi L, Sorbi S. Brain-derived neurotrophic factor genetic variants are not susceptibility factors to Alzheimer's disease in Italy. *Ann Neurol*. 2004; 55:447–448. [PubMed: 14991827]
  3. Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet*. 2010; 11:773–785. [PubMed: 20940738]
  4. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*. 1995; 57:289–300.
  5. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*. 2001; 29:1165–1188.
  6. Benner A, Zucknick M, Hielscher T, Itrich C, Mansmann U. High-Dimensional Cox Models: The Choice of Penalty as Part of the Model Building Process. *Biometrical Journal*. 2010; 52:50–69. [PubMed: 20166132]
  7. Bovelstad HM, Nygard S, Storvold HL, Aldrin M, Borgan O, Frigessi A, Lingjaerde OC. Predicting survival from microarray data - a comparative study. *Bioinformatics*. 2007; 23:2080–2087. [PubMed: 17553857]
  8. Braskie M, Ringman J, Thompson PM. Neuroimaging measures as endophenotypes in Alzheimer's disease. *International Journal of Alzheimer's Disease*. 2011 Feb. [in press].
  9. Brown CJ, Lain S, Verma CS, Fersht AR, Lane DP. Awakening guardian angels: drugging the p53 pathway. *Nat Rev Cancer*. 2009; 9:862–873. [PubMed: 19935675]
  10. Cannon TD, Keller MC. Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol*. 2006; 2:267–290. [PubMed: 17716071]
  11. Chapman J, Whittaker J. Analysis of multiple SNPs in a candidate gene or region. *Genet Epidemiol*. 2008; 32:560–566. [PubMed: 18428428]
  12. Chapuis J, Hannequin D, Pasquier F, Benthay P, Brice A, Leber I, Frebourg T, Deleuze JF, Cousin E, Thaker U, Amouyel P, Mann D, Lendon C, Campion D, Lambert JC. Association study of the GAB2 gene with the risk of developing Alzheimer's disease. *Neurobiol Dis*. 2008; 30:103–106. [PubMed: 18272374]

13. Chiang MC, Barysheva M, Toga AW, Medland SE, Hansell NK, James MR, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. BDNF gene effects on brain circuitry replicated in 455 twins. *Neuroimage*. 2010 [epub ahead of print].
14. Chiang, MC.; Barysheva, M.; McMahon, KL.; de Zubicaray, GI.; Johnson, K.; Martin, NG.; Toga, AW.; Wright, MJ.; Thompson, PM. Understanding the Network Topology of Gene Action on Brain Microstructure: An N=531 Twin Study. Organization for Human Brain Mapping Conference; June 2011; 2011.
15. Chiang, MC.; McMahon, KL.; de Zubicaray, GI.; Martin, NG.; Toga, AW.; Wright, MJ.; Thompson, PM. Hierarchical Clustering of the Genetic Connectivity Matrix Reveals the Network Topology of Gene Action on Brain Microstructure. International Symposium of Biomedical Imaging; April 2011; 2011. in press
16. Combarros O, Infante J, Llorca J, Berciano J. Polymorphism at codon 66 of the brain-derived neurotrophic factor gene is not associated with sporadic Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2004; 18:55–58. [PubMed: 15084795]
17. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet*. 2009; 10:392–404. [PubMed: 19434077]
18. de Leon MJ, George AE, Reisberg B, Ferris SH, Kluger A, Stylopoulos LA, Miller JD, La Regina ME, Chen C, Cohen J. Alzheimer's disease: longitudinal CT studies of ventricular change. *AJR Am J Roentgenol*. 1989; 152:1257–1262. [PubMed: 2785749]
19. De Luca V, Wang H, Squassina A, Wong GWH, Yeomans J, Kennedy JL. Linkage of M5 muscarinic and alpha 7-nicotinic receptor genes on 15q13 to schizophrenia. *Neuropsychobiology*. 2004; 50:124–127. [PubMed: 15292665]
20. del Arco A, Satrustegui J. Molecular cloning of Aralar, a new member of the mitochondrial carrier superfamily that binds calcium and is present in human muscle and brain. *Journal of Biological Chemistry*. 1998; 273:23327–23334. [PubMed: 9722566]
21. Desai P, Nebes R, DeKosky ST, Kamboh MI. Investigation of the effect of brain-derived neurotrophic factor (BDNF) polymorphisms on the risk of late-onset Alzheimer's disease (AD) and quantitative measures of AD progression. *Neuroscience Letters*. 2005; 379:229–234. [PubMed: 15843069]
22. ENIGMA Consortium. Genome-Wide Association Meta-Analysis of Hippocampal Volume: Results from the ENIGMA Consortium. Organization for Human Brain Mapping Conference; June 2011; 2011.
23. Ewens, WJ.; Grant, G. Statistical methods in bioinformatics: an introduction. Springer; New York: 2001.
24. Fillard P, Arsigny V, Pennec X, Hayashi KM, Thompson PM, Ayache N. Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *Neuroimage*. 2007; 34:639–650. [PubMed: 17113311]
25. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33:341–355. [PubMed: 11832223]
26. Flint, J.; Greenspan, RJ.; Kendler, KS. How genes influence behavior. Oxford University Press; Oxford; New York: 2010.
27. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D,

Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007; 449:851–861. [PubMed: 17943122]

28. Hayasaka S, Nichols TE. Combining voxel intensity and cluster extent with permutation test framework. *Neuroimage*. 2004; 23:54–63. [PubMed: 15325352]
29. Hemminger BM, Saelim B, Sullivan PF. TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. *Bioinformatics*. 2006; 22:626–627. [PubMed: 16418238]
30. Hibar, D.; Stein, J.L.; Jahanshad, N.; Baresheva, M.; Feng, A.; Kogachi, S.; McMahon, K.; De Zubicaray, G.; Hansell, N.; Martin, N.G.; Wright, M.J.; Toga, A.; Thompson, P. Voxelwise genome-wide association of Diffusion Tensor Images identifies putative novel variants influencing white matter integrity in 467 related young adults. *Society for Neuroscience*; San Diego, CA: 2010.
31. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. 2006; 34:D590–598. [PubMed: 16381938]
32. Hirabayashi S, Tajima M, Yao I, Nishimura W, Mori H, Hata Y. JAM4, a junctional cell adhesion molecule interacting with a tight junction protein, MAGI-1. *Molecular and Cellular Biology*. 2003; 23:4267–4282. [PubMed: 12773569]
33. Ho AJ, Hua X, Lee S, Leow AD, Yanovsky I, Gutman B, Dinov ID, Lepore N, Stein JL, Toga AW, Jack CR, Bernstein MA, Reiman EM, Harvey DJ, Kornak J, Schuff N, Alexander GE, Weiner MW, Thompson PM. Neuroimaging, A.s.D. Comparing 3 T and 1.5 T MRI for Tracking Alzheimer's Disease Progression with Tensor-Based Morphometry. *Human Brain Mapping*. 2010; 31:499–514. [PubMed: 19780044]
34. Hua X, Leow AD, Parikshak N, Lee S, Chiang MC, Toga AW, Jack CR, Weiner MW, Thompson PM. Initi, A.s.D.N. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: An MRI study of 676 AD, MCI, and normal subjects. *Neuroimage*. 2008; 43:458–469. [PubMed: 18691658]
35. Ikram MA, Liu F, Oostra BA, Hofman A, van Duijn CM, Breteler MMB. The GAB2 Gene and the Risk of Alzheimer's Disease: Replication and Meta-Analysis. *Biological Psychiatry*. 2009; 65:995–999. [PubMed: 19118819]
36. Inkster B, Nichols TE, Saemann PG, Auer DP, Holsboer F, Muglia P, Matthews PM. Pathway-based approaches to imaging genetics association studies: Wnt signaling, GSK3beta substrates and major depression. *Neuroimage*. 2010; 53:908–917. [PubMed: 20219685]
37. International Hap Map Consortium. A haplotype map of the human genome. *Nature*. 2005; 437:1299–1320. [PubMed: 16255080]
38. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004; 431:931–945. [PubMed: 15496913]

39. Ishiguro H, Gong JP, Hall FS, Arinami T, Uhl GR. Association of PTPRB Gene Polymorphism With Drug Addiction. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*. 2008; 147B:1167–1172.
40. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DLG, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW, Study A. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*. 2008; 27:685–691. [PubMed: 18302232]
41. Jagannathan K, Calhoun VD, Gelernter J, Stevens MC, Liu J, Bolognani F, Windemuth A, Ruaño G, Assaf M, Pearlson GD. Genetic Associations of Brain Structural Networks in Schizophrenia: A Preliminary Study. *Biological Psychiatry*. 2010; 68:657–666. [PubMed: 20691427]
42. Jovicich J, Czanner S, Greve D, Haley E, van der Kouwe A, Gollub R, Kennedy D, Schmitt F, Brown G, MacFall J, Fischl B, Dale A. Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *Neuroimage*. 2006; 30:436–443. [PubMed: 16300968]
43. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient control of population structure in model organism association mapping. *Genetics*. 2008; 178:1709–1723. [PubMed: 18385116]
44. Kleinbaum, DG. *Applied regression analysis and other multivariable methods*. 4. Brooks/Cole; Australia; Belmont, CA: 2007.
45. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994; 265:2037–2048. [PubMed: 8091226]
46. Langers DR, Jansen JF, Backes WH. Enhanced signal detection in neuroimaging by means of regional control of the global false discovery rate. *Neuroimage*. 2007; 38:43–56. [PubMed: 17825583]
47. Leow A, Huang SC, Geng A, Becker J, Davis S, Toga A, Thompson P. Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties. *Inf Process Med Imaging*. 2005; 19:493–503. [PubMed: 17354720]
48. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet*. 2009; 10:387–406. [PubMed: 19715440]
49. Lin K, Tang M, Han H, Guo Y, Lin Y, Ma C. GAB2 is not associated with late-onset Alzheimer's disease in Chinese Han. *Neurol Sci*. 2010; 31:277–281. [PubMed: 19924507]
50. Liu J, Pearlson G, Windemuth A, Ruano G, Perrone-Bizzozero NI, Calhoun V. Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Hum Brain Mapp*. 2009; 30:241–255. [PubMed: 18072279]
51. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 2010; 87:139–145. [PubMed: 20598278]
52. Lucek P, Hanke J, Reich J, Solla SA, Ott J. Multi-locus nonparametric linkage analysis of complex trait loci with neural networks. *Hum Hered*. 1998; 48:275–284. [PubMed: 9748698]
53. Malo N, Libiger O, Schork NJ. Accommodating linkage disequilibrium in genetic-association analyses via ridge regression. *Am J Hum Genet*. 2008; 82:375–385. [PubMed: 18252218]
54. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet*. 2005; 37:413–417. [PubMed: 15793588]
55. Matsushita S, Arai H, Matsui T, Yuzuriha T, Urakami K, Masaki T, Higuchi S. Brain-derived neurotrophic factor gene polymorphisms and Alzheimer's disease. *J Neural Transm*. 2005; 112:703–711. [PubMed: 15375678]
56. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci*. 2001; 356:1293–1322. [PubMed: 11545704]



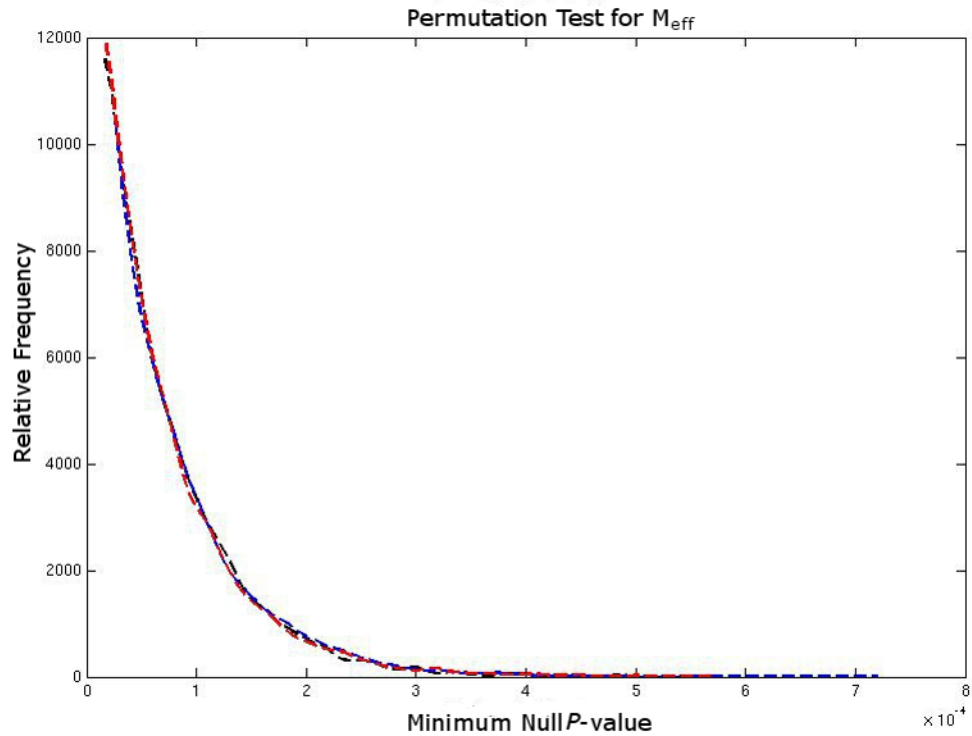
57. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008; 9:356–369. [PubMed: 18398418]
58. Mukhopadhyay I, Feingold E, Weeks DE, Thalamuthu A. Association Tests Using Kernel-Based Measures of Multi-Locus Genotype Similarity Between Individuals. *Genetic Epidemiology.* 2010; 34:213–221. [PubMed: 19697357]
59. Nacmias B, Piccini C, Bagnoli S, Tedde A, Cellini E, Bracco L, Sorbi S. Brain-derived neurotrophic factor, apolipoprotein E genetic variants and cognitive performance in Alzheimer's disease. *Neuroscience Letters.* 2004; 367:379–383. [PubMed: 15337270]
60. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. *Am J Hum Genet.* 2004; 75:353–362. [PubMed: 15272419]
61. Nishimura AL, Oliveira JR, Mitne-Neto M, Guindalini C, Nitrini R, Bahia VS, de Brito-Marques PR, Otto PA, Zatz M. Lack of association between the brain-derived neurotrophin factor (C-270T) polymorphism and late-onset Alzheimer's disease (LOAD) in Brazilian patients. *J Mol Neurosci.* 2004; 22:257–260. [PubMed: 14997020]
62. Ott J. Neural networks and disease association studies. *Am J Med Genet.* 2001; 105:60–61. [PubMed: 11425001]
63. Parsons CG, Stoffler A, Danysz W. Memantine: a NMDA receptor antagonist that improves memory by restoration of homeostasis in the glutamatergic system--too little activation is bad, too much is even worse. *Neuropharmacology.* 2007; 53:699–723. [PubMed: 17904591]
64. Paus, T.; Bernard, M.; Chakravarty, M.; Lourdasamy, A.; Leonard, G.; Perron, M.; Pike, B.; Richer, L.; Schumann, G.; Veillette, S.; Pausova, Z. Association between KCTD8 and Brain Volume as Revealed in a Genome-wide Study. *Organization for Human Brain Mapping Conference*; June 2011; 2011.
65. Petersen RC. Aging, mild cognitive impairment, and Alzheimer's disease. *Neurol Clin.* 2000; 18:789–806. [PubMed: 11072261]
66. Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, Saykin AJ, Orro A, Lupoli S, Salvi E, Weiner M, Macciardi F. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. *PLoS One.* 2009; 4:e6501. [PubMed: 19668339]
67. Potkin SG, Turner JA, Guffanti G, Lakatos A, Fallon JH, Nguyen DD, Mathalon D, Ford J, Lauriello J, Macciardi F. A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophr Bull.* 2009; 35:96–108. [PubMed: 19023125]
68. Potkin SG, Turner JA, Guffanti G, Lakatos A, Torri F, Keator DB, Macciardi F. Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. *Cogn Neuropsychiatry.* 2009; 14:391–418. [PubMed: 19634037]
69. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]
70. Rabionet R, Jaworski JM, Ashley-Koch AE, Martin ER, Sutcliffe JS, Haines JL, DeLong GR, Abramson RK, Wright HH, Cuccaro ML, Gilbert JR, Pericak-Vance MA. Analysis of the autism chromosome 2 linkage region: GAD1 and other candidate genes. *Neuroscience Letters.* 2004; 372:209–214. [PubMed: 15542242]
71. Rajagopalan, P.; Jahanshad, N.; Chiang, MC.; Stein, JL.; Hibar, DP.; Ryles, A.; McMahon, KL.; de Zubicaray, GI.; Martin, NM.; Wright, MJ.; Saykin, AJ.; Jack, CR., Jr; Weiner, MW.; Toga, AW.; Thompson, PM. and the Alzheimer's Disease Neuroimaging Initiative. Folate gene variant is associated with brain volume differences: Replication in ADNI (N=740) and Queensland Twins (N=577). *Organization for Human Brain Mapping Conference*; June 2011; 2011.
72. Ramirez-Lorca R, Boada M, Saez ME, Hernandez I, Mauleon A, Rosende-Roca M, Martinez-Lage P, Gutierrez M, Real LM, Lopez-Arrieta J, Gayan J, Antunez C, Gonzalez-Perez A, Tarraga L, Ruiz A. GAB2 gene does not modify the risk of Alzheimer's disease in Spanish APOE 4 carriers. *J Nutr Health Aging.* 2009; 13:214–219. [PubMed: 19262956]

73. Ramoz N, Reichert JG, Smith CJ, Silverman JM, Beshpalova IN, Davis KL, Buxbaum JD. Linkage and association of the mitochondrial aspartate/glutamate carrier SLC25A12 gene with autism. *Am J Psychiatry*. 2004; 161:662–669. [PubMed: 15056512]
74. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet*. 2001; 17:502–510. [PubMed: 11525833]
75. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ, Craig DW, Coon KD, Liang WS, Herbert RH, Beach T, Rohrer KC, Zhao AS, Leung D, Bryden L, Marlowe L, Kaleem M, Mastroeni D, Grover A, Heward CB, Ravid R, Rogers J, Hutton ML, Melquist S, Petersen RC, Alexander GE, Caselli RJ, Kukull W, Papassotiropoulos A, Stephan DA. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron*. 2007; 54:713–720. [PubMed: 17553421]
76. Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S, Risacher SL, Nho K, Huentelman MJ, Craig DW, Thompson PM, Stein JL, Moore JH, Farrer LA, Green RC, Bertram L, Jack CR Jr, Weiner MW. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimers Dement*. 2010; 6:265–273. [PubMed: 20451875]
77. Schaid DJ. Evaluating associations of haplotypes with traits. *Genet Epidemiol*. 2004; 27:348–364. [PubMed: 15543638]
78. Schjeide BM, Hooli B, Parkinson M, Hogan MF, DiVito J, Mullin K, Blacker D, Tanzi RE, Bertram L. GAB2 as an Alzheimer disease susceptibility gene: follow-up of genome-wide association results. *Arch Neurol*. 2009; 66:250–254. [PubMed: 19204163]
79. Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, Foroud T, Pankratz N, Moore JH, Sloan CD, Huentelman MJ, Craig DW, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Saykin AJ. 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; 53:1051–1063. [PubMed: 20100581]
80. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*. 1998; 17:87–97. [PubMed: 9617910]
81. Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, DeChairo BM, Potkin SG, Weiner MW, Thompson P. Voxelwise genome-wide association study (vGWAS). *Neuroimage*. 2010; 53:1160–1174. [PubMed: 20171287]
82. Stein JL, Hua X, Morra JH, Lee S, Hibar DP, Ho AJ, Leow AD, Toga AW, Sul JH, Kang HM, Eskin E, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Stephan DA, Webster J, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Thompson PM. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. *Neuroimage*. 2010; 51:542–554. [PubMed: 20197096]
83. Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. *Annals of Statistics*. 2003; 31:2013–2035.
84. Sun YV, Shedden KA, Zhu J, Choi NH, Kardia SL. Identification of correlated genetic variants jointly associated with rheumatoid arthritis using ridge regression. *BMC Proc*. 2009; 3(Suppl 7):S67. [PubMed: 20018061]
85. Telliez JB, Bean KM, Lin LL. LRDD, a novel leucine rich repeat and death domain containing protein. *Biochim Biophys Acta*. 2000; 1478:280–288. [PubMed: 10825539]
86. Thode, HC. Testing for normality. Marcel Dekker; New York: 2002.
87. Thompson PM, Martin NG, Wright MJ. Imaging genomics. *Curr Opin Neurol*. 2010; 23:368–373. [PubMed: 20581684]
88. Tsai SJ, Hong CJ, Liu HC, Liu TY, Hsu LE, Lin CH. Association analysis of brain-derived neurotrophic factor Val66Met polymorphisms with Alzheimer's disease and age of onset. *Neuropsychobiology*. 2004; 49:10–12. [PubMed: 14730194]
89. Vepsäläinen S, Castren E, Helisalmi S, Iivonen S, Mannermaa A, Lehtovirta M, Hanninen T, Soininen H, Hiltunen M. Genetic analysis of BDNF and TrkB gene polymorphisms in Alzheimer's disease. *J Neurol*. 2005; 252:423–428. [PubMed: 15726264]

90. Vounou M, Nichols TE, Montana G. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage*. 2010; 53:1147–1159. [PubMed: 20624472]
91. Wang K, Abbott D. A principal components regression approach to multilocus genetic association studies. *Genet Epidemiol*. 2008; 32:108–118. [PubMed: 17849491]
92. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet*. 2005; 6:109–118. [PubMed: 15716907]
93. Wang T, Elston RC. Improved power by use of a weighted score test for linkage disequilibrium mapping. *Am J Hum Genet*. 2007; 80:353–360. [PubMed: 17236140]
94. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–678. [PubMed: 17554300]
95. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Ostell J, Pruitt KD, Schuler GD, Shumway M, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E. Database resources of the national center for biotechnology information. *Nucleic Acids Research*. 2008; 36:D13–D21. [PubMed: 18045790]
96. Wu CW, Kao HL, Li AFY, Chi CW, Lin WC. Protein tyrosine-phosphatase expression profiling in gastric cancer tissues. *Cancer Letters*. 2006; 242:95–103. [PubMed: 16338072]
97. Wu TT, Chen YF, Hastie T, Sobel E, Lange K. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*. 2009; 25:714–721. [PubMed: 19176549]
98. Yu WH, Fraser PE. S100 beta interaction with tau is promoted by zinc and inhibited by hyperphosphorylation in Alzheimer's disease. *Journal of Neuroscience*. 2001; 21:2240–2246. [PubMed: 11264299]
99. Zou H. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*. 2006; 101:1418–1429.

**Research Highlights**

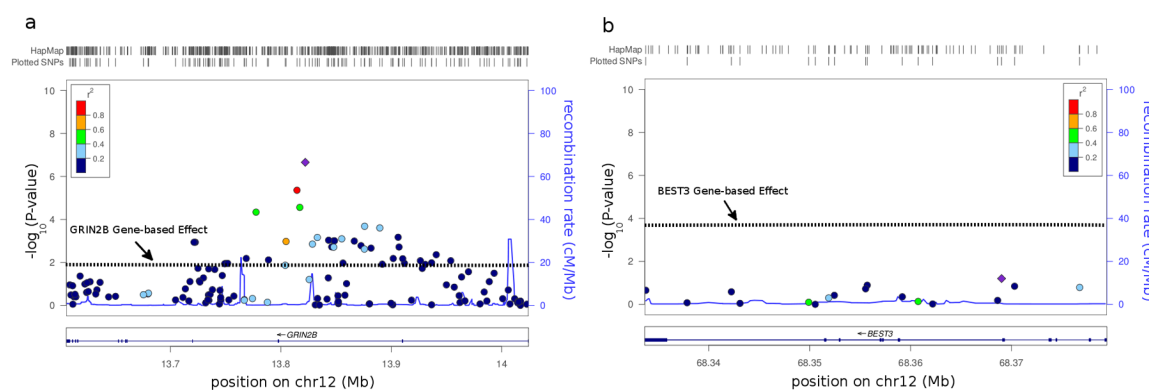
- Principal components regression for gene-based association models
- Gene-based multivariate statistics have increased power over univariate methods
- Gene-based testing is a useful complement to genome-wide association scans



**Figure 1.**

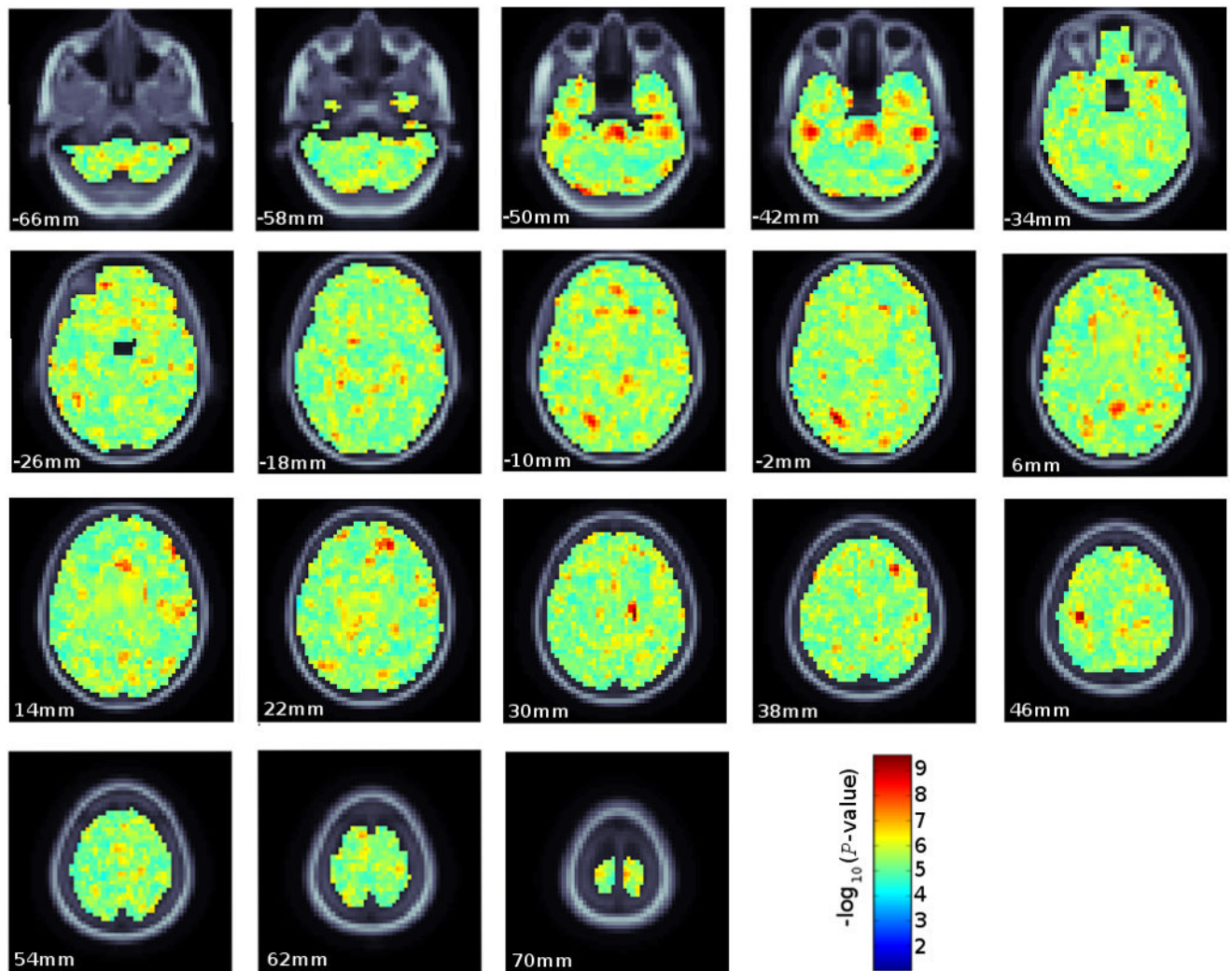
A histogram shows the minimum null  $P$ -values obtained from permutation tests. Data from 3 different voxels are shown on the same graph (blue, red, and black lines); they are obtained from 3 randomly chosen, uncorrelated voxels in the brain (5000 permutations each). The distributions are nearly identical, and agree with each other, as well as accurately reflecting the effective number of independent tests ( $M_{\text{eff}}$ ).





**Figure 2.**

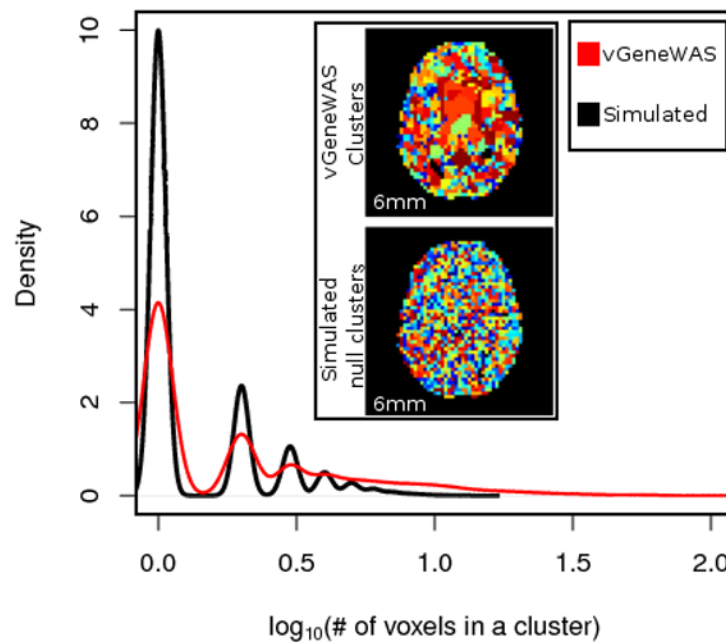
Genetic association plots for univariate linear regression versus multi-locus PCReg. The  $-\log_{10}(P\text{-value})$  of each SNP in *GRIN2B* (a) and *BEST3* (b) is plotted against its position in the gene. Each of the points is color coded by level of LD (compared to the top SNP, the purple diamond dot) as measured by  $r^2$ . The  $-\log_{10}(P\text{-value})$  of the gene-based PCReg test for each gene is overlaid on the plot for comparison (dotted black line). Plots were generated using the *LocusZoom* software (<http://csg.sph.umich.edu/locuszoom/>).



**Figure 3.**

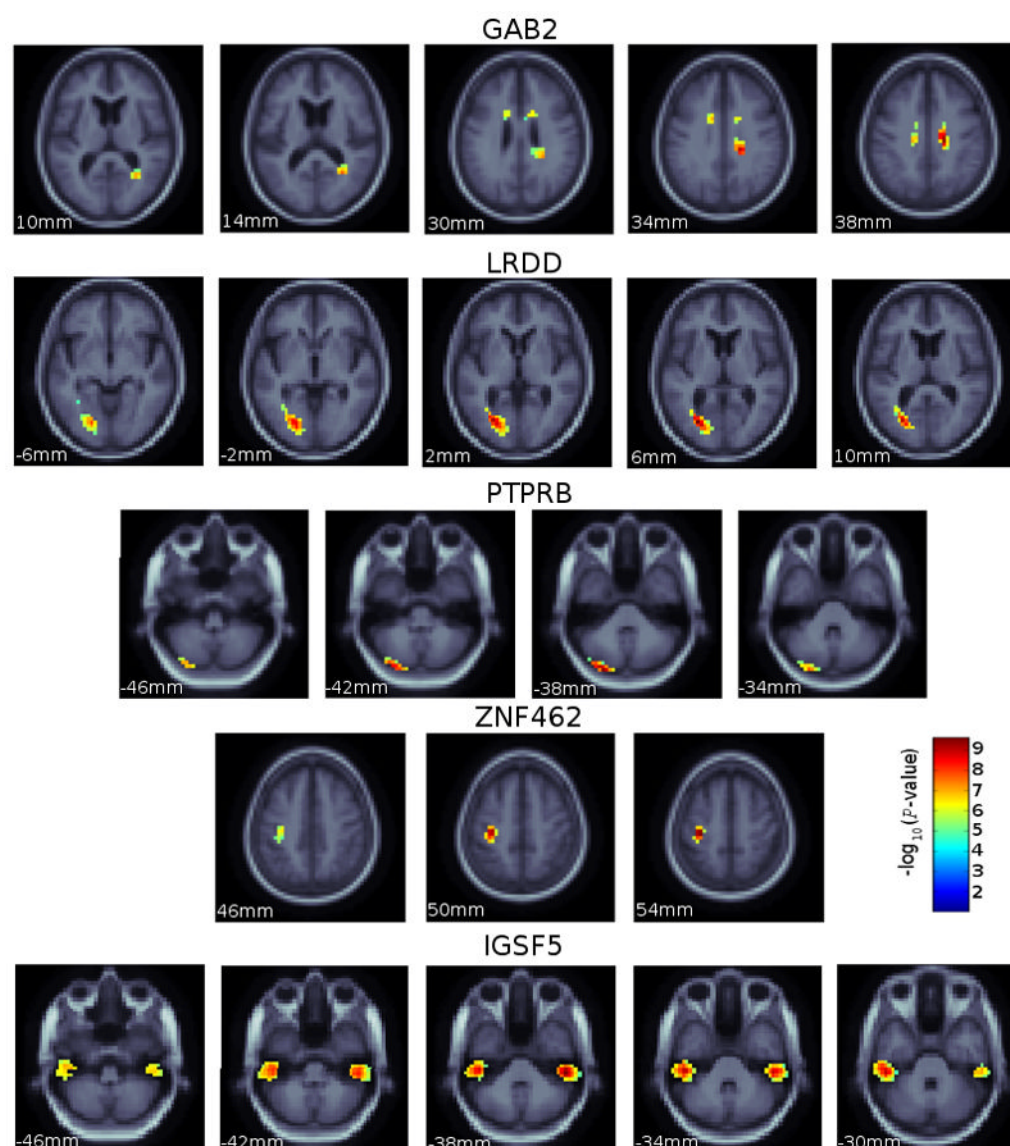
A color-coded significance map of the top gene at each voxel. Sections are shown at 8mm intervals throughout the brain. The top of each panel represents the anterior of the brain and bottom the posterior of the brain. The images are in radiological convention (the left side of the image is the patient's right hemisphere). Color coding is based on the  $-\log_{10}(P\text{-value})$ ; warmer colors represent more significant associations.

### Cluster Size Density Distribution Comparison

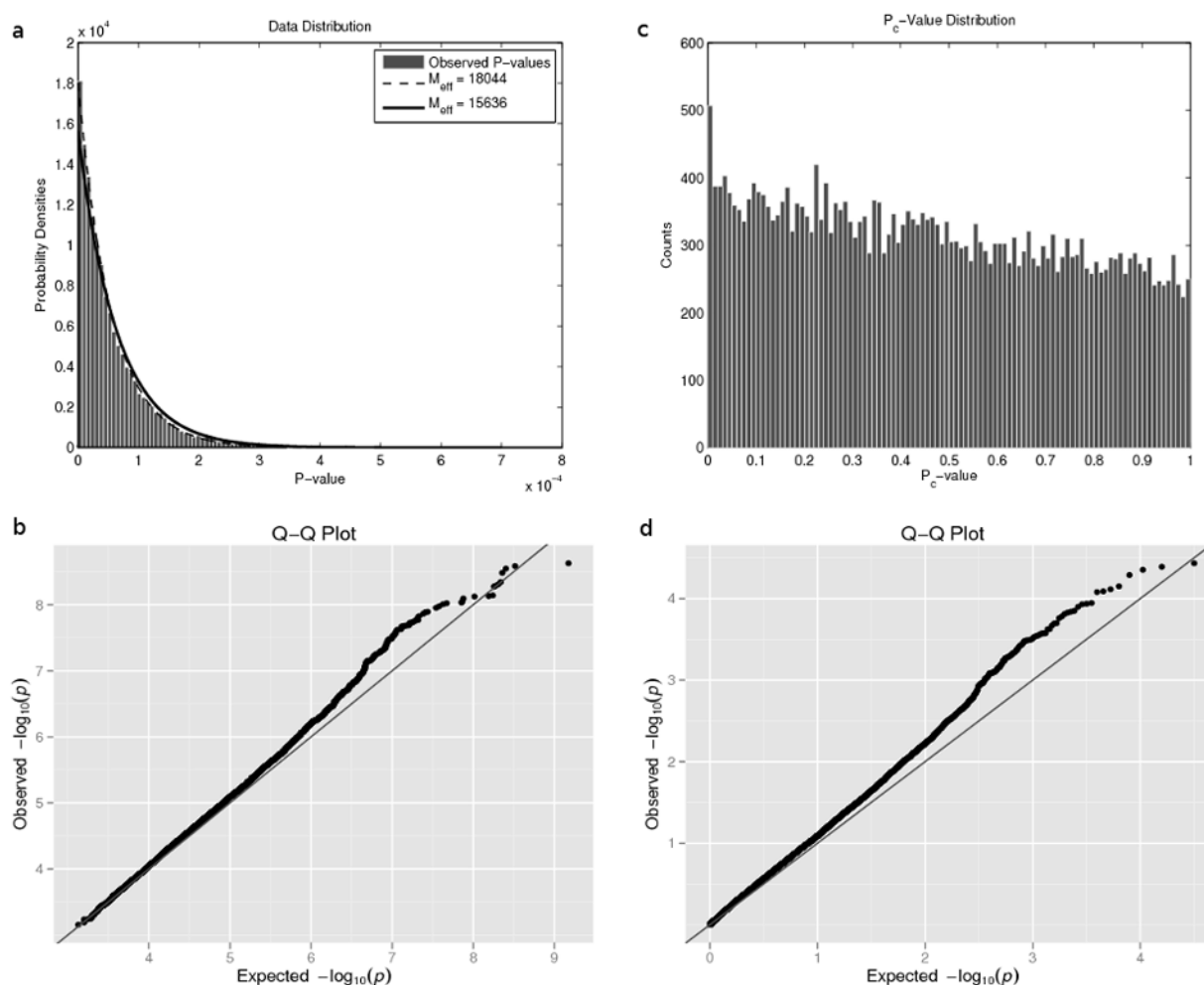


**Figure 4.**

Cluster sizes in vGeneWAS (*red line*) are compared with an average of simulated null maps (*black line*). We took the  $\log_{10}$  of the number of voxels in a cluster (not in  $\text{mm}^3$ ) across the brain in both maps for scaling purposes and for ease of comparison. The  $\log_{10}$  cluster sizes are then plotted using a density function such that the total area under each line is equal to 1. The average simulated null map contains a larger proportion of small cluster sizes than vGeneWAS (higher peaks in the black line at values close to the origin on the  $x$ -axis). The vGeneWAS map contains a larger proportion of large cluster sizes than the average simulated null map (the red line is higher at larger values and is more extended). A single slice view of the vGeneWAS and average simulated null cluster maps are pictured for comparison (*inset*). Every unique cluster is assigned its own color. There are more unique clusters than distinct colors making visual inspection difficult, but in general the clusters in the vGeneWAS maps are larger.



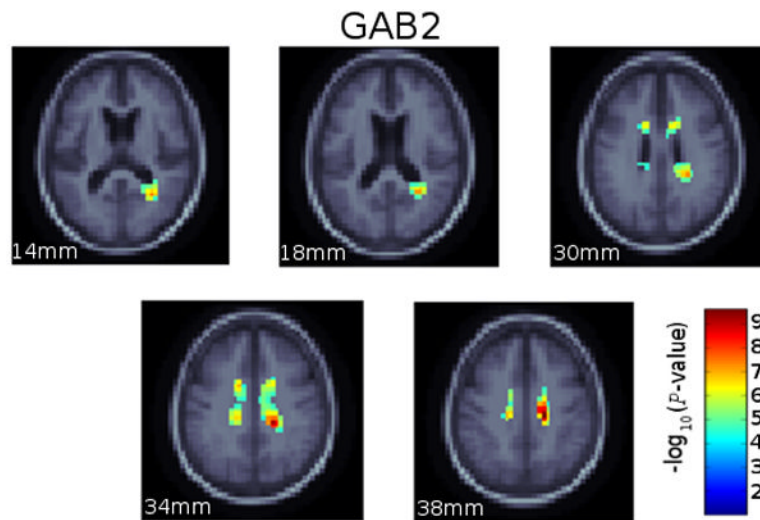
**Figure 5.** Regions in the brain associated with the top 5 genes from our vGeneWAS analysis (where the uncorrected  $P$ -value at a given voxel is overlaid on the minimum deformation template). The slices chosen best represent the regions where each gene was the most significantly associated gene in the brain. Images read from inferior to superior (left to right of the page) following radiological convention and with the top and bottom of each panel representing the anterior and posterior of the brain, respectively.



**Figure 6.**

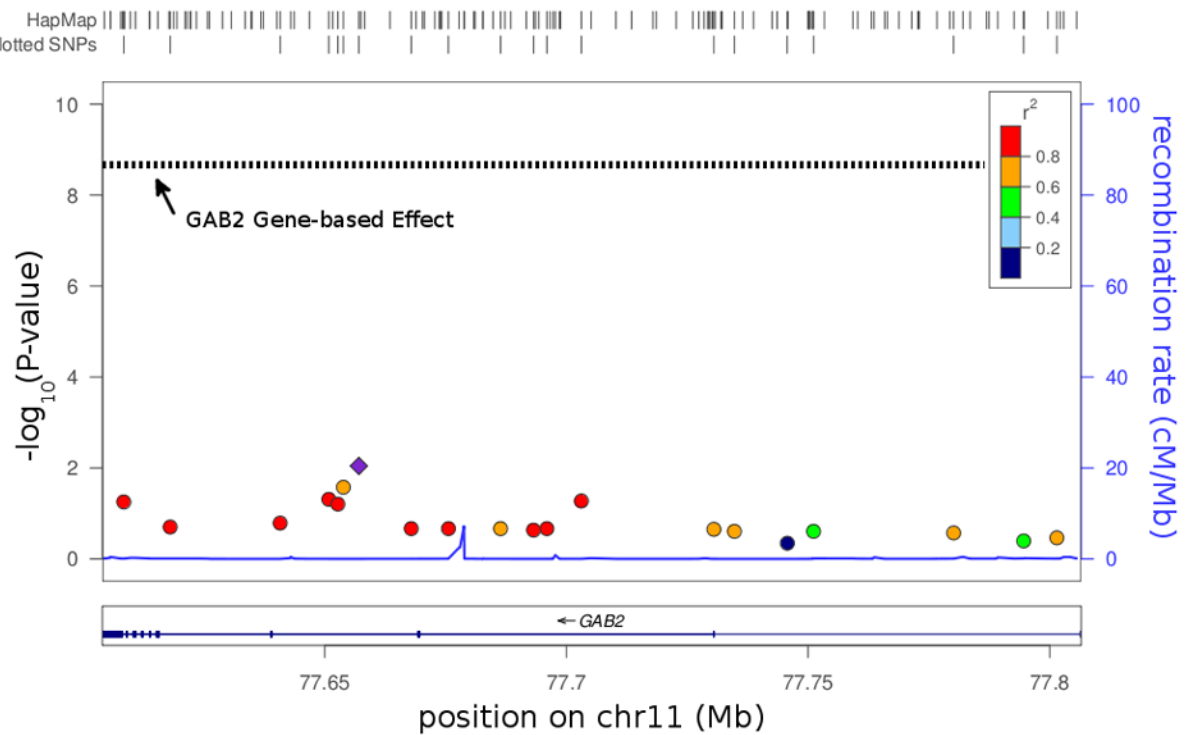
(a) The normalized histogram of observed  $P$ -values. The dashed line represents the cumulative distribution function (CDF) of Beta(1, 18044) where  $M_{\text{eff}}$  is based on the number of genes tested. The solid line represents the CDF of Beta(1, 15636) where  $M_{\text{eff}}$  is an estimate of the number of independent tests from permutation testing. (b) The Q-Q plot shows the observed  $P$ -values versus those expected from a Beta(1, 15636) distribution (black dots). The solid gray line represents a purely null distribution of  $P$ -values. (c) The histogram of corrected  $P$ -values ( $P_c$ ) approximately follows a uniform distribution. (d) The Q-Q plot of the observed  $P_c$  versus those expected from a null distribution.





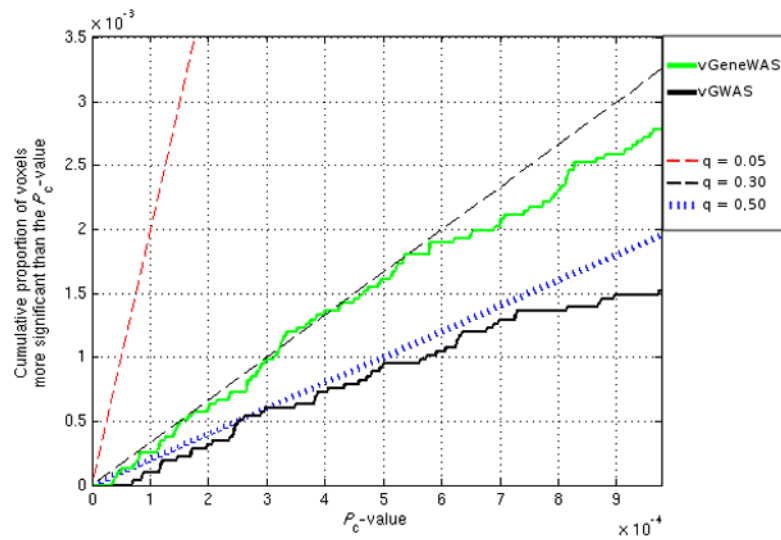
**Figure 7.**

Map of  $P$ -values for  $GAB2$  at every voxel in the brain after correction for multiple comparisons across voxels (but not corrected for search across the genome, as we are only testing one gene) using the original FDR method.  $P$ -values significant after FDR correction (at  $q = 3.4 \times 10^{-4}$ ) are color-coded. Warmer colors are more significant.  $GAB2$  has a more distributed effect on the brain than is evident in the vGeneWAS results.



**Figure 8.**

Genetic association plot, for different SNPs in the *GAB2* gene, at the top voxel from our analysis. The  $-\log_{10}(P\text{-value})$  of each SNP in *GAB2* is plotted against its position in the gene. Each of the points is color coded by level of LD (compared to the top SNP, the purple diamond dot) as measured by  $r^2$ . The  $-\log_{10}(P\text{-value})$  of the gene-based PCReg test for *GAB2* at this voxel is overlaid on the plot for comparison (dotted black line). In this case, the gene-based test shows a greater effect size than univariate tests on any of the component SNPs treated independently. This shows that the gene-based test can be more powerful than performing separate tests on component SNPs.



**Figure 9.**

vGeneWAS may control the false discovery rate better than vGWAS. The cumulative distribution function (CDF) of  $P_c$ -values from vGeneWAS (*solid green line*) is compared to the CDF of  $P_c$ -values from vGWAS (Stein et al. 2010a; *solid black line*). Three lines represent different correction thresholds of  $q = 0.05$  (*red dashed line*),  $q = 0.30$  (*black dashed line*), and  $q = 0.50$  (*blue dotted line*).

The top 20 genes most significantly associated at any voxel, organized by minimum observed  $P$ -value. The common gene name is listed, with the number of SNPs within that gene (Note: the number of SNPs in a gene will vary depending on the genotyping and quality control methods used). We also list the number of eigenSNPs included in the PCReg models. The mean observed  $P$ -value is also listed; it is the average of the  $P$ -values at all voxels where that gene was most significantly associated. Also listed is the volume of voxels (in  $\text{mm}^3$ ) and the percentage of the total volume of voxels in the brain where the gene was the most significantly associated gene. The maximum cluster size observed for each gene is listed as well as the total number of spatially independent clusters.

**Table 1**

Chr	Gene	# of SNPs in Gene	# of eigenSN Ps	Minimum P-value	Mean P-value	Volume ( $\text{mm}^3$ )	% of brain volume	Cluster <sub>max</sub> ( $\text{mm}^3$ )	# of Clusters
11	GAB2	20	10	$2.36 \times 10^{-9}$	$1.50 \times 10^{-5}$	6336	0.0049	2688	9
11	LRDD	2	2	$2.60 \times 10^{-9}$	$1.32 \times 10^{-5}$	8128	0.0063	7872	4
12	PTPRB	17	13	$2.84 \times 10^{-9}$	$1.81 \times 10^{-5}$	3200	0.0024	3008	5
9	ZNF462	9	6	$3.29 \times 10^{-9}$	$1.84 \times 10^{-5}$	2688	0.0021	2688	1
21	IGSF5	27	14	$5.32 \times 10^{-9}$	$1.62 \times 10^{-5}$	16384	0.013	9344	3
2	SLC25A12	10	5	$9.48 \times 10^{-9}$	$2.66 \times 10^{-5}$	1856	0.0014	1792	2
11	MRE11A	9	6	$9.86 \times 10^{-9}$	$8.80 \times 10^{-6}$	9344	0.0072	9216	3
19	SLC8A2	11	7	$1.06 \times 10^{-8}$	$3.18 \times 10^{-5}$	5632	0.0043	5376	3
15	CHRM5	3	3	$1.71 \times 10^{-8}$	$1.77 \times 10^{-5}$	1280	0.00099	1216	2
18	SPIRE1	19	14	$2.94 \times 10^{-8}$	$2.88 \times 10^{-5}$	6016	0.0046	3072	12
3	C3orf64	9	8	$3.71 \times 10^{-8}$	$2.43 \times 10^{-5}$	4352	0.0034	2112	4
21	S100B	1	1	$4.75 \times 10^{-8}$	$2.81 \times 10^{-5}$	9344	0.0072	6656	7
1	CRCT1	1	1	$5.54 \times 10^{-8}$	$2.90 \times 10^{-5}$	4096	0.0032	3456	4
19	ZNF626	6	5	$5.85 \times 10^{-8}$	$2.12 \times 10^{-5}$	2560	0.0020	2112	3
1	ELK4	1	1	$6.05 \times 10^{-8}$	$3.27 \times 10^{-5}$	4032	0.0031	2688	4
11	RSF1	8	6	$9.30 \times 10^{-8}$	$1.31 \times 10^{-5}$	768	0.00059	768	1
20	WFDC11	2	2	$1.06 \times 10^{-7}$	$2.49 \times 10^{-5}$	1280	0.00099	512	5
6	SCML4	27	18	$1.07 \times 10^{-7}$	$1.67 \times 10^{-5}$	3328	0.0026	3328	1
12	ERP27	8	14	$1.08 \times 10^{-7}$	$2.61 \times 10^{-5}$	2624	0.0020	2176	2