NeuroImage 159 (2017) 159–169

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Imaging-wide association study: Integrating imaging endophenotypes in GWAS

Zhiyuan Xu, Chong Wu, Wei Pan *, for the Alzheimer's Disease Neuroimaging Initiative[1]

Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

Keywords:
Alzheimer's disease
aSPU test
IWAS
MRI
SPU test
SSU test
Sum test
TWAS

ABSTRACT

A new and powerful approach, called imaging-wide association study (IWAS), is proposed to integrate imaging endophenotypes with GWAS to boost statistical power and enhance biological interpretation for GWAS discoveries. IWAS extends the promising transcriptome-wide association study (TWAS) from using gene expression endophenotypes to using imaging and other endophenotypes with a much wider range of possible applications. As illustration, we use gray-matter volumes of several brain regions of interest (ROIs) drawn from the ADNI-1 structural MRI data as imaging endophenotypes, which are then applied to the individual-level GWAS data of ADNI-GO/2 and a large meta-analyzed GWAS summary statistics dataset (based on about 74,000 individuals), uncovering some novel genes significantly associated with Alzheimer's disease (AD). We also compare the performance of IWAS with TWAS, showing much larger numbers of significant AD-associated genes discovered by IWAS, presumably due to the stronger link between brain atrophy and AD than that between gene expression of normal individuals and the risk for AD. The proposed IWAS is general and can be applied to other imaging endophenotypes, and GWAS individual-level or summary association data.

1. Introduction

During the last fifteen years, genome-wide association studies (GWAS) have been quite successful in identifying thousands of risk loci associated with complex diseases and traits. Due to linkage disequilibrium (LD), most time it is difficult to pinpoint causal genetic variants or genes, thus mechanistic interpretations of underlying biology remain largely elusive. Furthermore, the uncovered risk loci only account for a small proportion of the heritability for each complex trait. For example, for Alzheimer's disease (AD), multiple common and rare variants have been identified (Marei et al., 2016; Saykin et al., 2015), among which the APOEε4 allele has been consistently shown to be associated with AD (Lambert et al., 2009). However, only 50% of AD patients carry an APOEε4 allele, suggesting the existence of other genetic variants contributing to risk for the disease (Karch et al., 2014). A recent study indicates that 33% of total AD phenotypic variance is explained by common variants; APOE alone explains 6% and other known markers 2%, meaning more than 25% of phenotypic variance remains unexplained by known common variants (Ridge et al., 2013). Hence, as for other common and complex diseases and traits, many more genetic factors underlying late onset AD remain to be discovered (Bertram and Tanzi, 2011). One obvious but costly approach is to have a larger sample size, motivating meta- or mega-analyses by large consortia. At the same time, a complementary strategy is to use multiple endophenotypes, intermediate between genetics and the disease (Gottesman and Gould, 2003; Glahn et al., 2012). The potential of the strategy has been demonstrated by a recent GWAS: some risk genes, such as *FRMD6*, were first identified to be associated with some neuroimaging endophenotypes (e.g. hippocampal atrophy) (Shen et al., 2014), then were later validated to be associated with AD (Hong et al., 2012; Sherva et al., 2014).

However, existing methods use imaging endophenotypes as GWAS traits to directly identify endophenotypes-associated single nucleotide polymorphisms (SNPs) or genes (e.g., Lin et al., 2014; Shen et al., 2014; Zhu et al., 2014; Huang et al., 2015; Lu et al., 2017; Tao et al., 2017, and references therein). Since the identified SNPs or genes may or may not be associated with the disease, e.g. AD, further studies are still needed to confirm or refute a suggestive link based on imaging endophenotypes. In addition, due to the high cost and logistic difficulties, despite recent

* Corresponding author.
E-mail address: weip@biostat.umn.edu (W. Pan).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). 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. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

http://dx.doi.org/10.1016/j.neuroimage.2017.07.036
Received 2 May 2017; Received in revised form 22 June 2017; Accepted 18 July 2017
Available online 20 July 2017

efforts in forming large consortia for imaging studies (Thompsona et al., 2013), the sample size of a typical GWAS with imaging traits is still much smaller than those of other GWAS with clinical traits, hindering the discovery of SNPs associated with imaging endophenotypes, as shown by the ADNI data (Shen et al., 2014; Saykin et al., 2015). Alternatively, we extend the idea of *transcriptome-wide association study (TWAS)* (Gamazon et al., 2015; Gusev et al., 2016) to *imaging-wide association study (IWAS)*: instead of using gene expression as an endophenotype, we use an imaging endophenotype to construct weights for a weighted gene-based GWAS test. TWAS is motivated by possible regulatory roles of eQTL (expression quantitative trait loci) or expression SNPs which are more likely to be disease-associated (Nicolae et al., 2010). Accordingly, instead of directly associating observed gene expression levels with a GWAS trait, TWAS considers the *genetically regulated component* of gene expression and its association with the GWAS trait. Analogously, we consider the *genetically regulated component* of an imaging endophenotype and its association with the GWAS trait; the genetically regulated component of an imaging endophenotype is not only directly related to the goal of genetic association analysis, but also excludes the noise of an observed endophenotype, which for example is likely to contain a component influenced by various environmental factors. We take advantage of possible link between AD (or other neuro-degenerative diseases) and brain atrophy as reflected by neuroimaging features, which is treated as prior knowledge and incorporated into weighted gene-based testing in GWAS. A possibly useful but under-utilized endophenotype is the brain default mode network (DMN), consisting of several brain regions of interest (ROIs) remaining active in the resting state. Brain activity in DMN may explain the etiology of AD (Metin et al., 2015), and is a plausible indicator for incipient AD (Damoiseaux et al., 2012; Greicius et al., 2004; He et al., 2009; Jones et al., 2011; Balthazar et al., 2014). Since there is growing evidence that genetic factors play a role in aberrant default mode connectivity (Glahn et al., 2010), it will be more powerful to detect genetic variants associated with AD by taking advantage of DMN as imaging endophenotypes.

Since there may be multiple imaging endophenotypes related to both genetic factors and a complex trait, we may consider each endophenotype separately before combining them together. As to be shown in our example, the gray matter volume in each of 12 ROIs related to DMN will be considered as an endophenotype in an integrated GWAS analysis to detect SNPs associated with AD. Hence, we call our approach as *imaging-wide association study* (IWAS). Similar to TWAS, in addition to a *main* GWAS dataset with a trait of interest (e.g., AD), we require another independent *reference* GWAS dataset with imaging endophenotypes; as for a reference eQTL dataset in TWAS (Gusev et al., 2016), the required sample size of a reference GWAS dataset in IWAS can be much smaller than that of the main GWAS dataset. An advantage of IWAS is the increasing availability of such data. For example, *Alzheimer's Disease Neuroimaging Initiative (ADNI)* has published such reference GWAS data with both GWAS genotypes and a wide-range of imaging endophenotypes (Shen et al., 2014). Furthermore, in addition to main GWAS individual-level genotypic and phenotypic data, IWAS is also applicable to GWAS summary statistics from single GWAS or meta-analyses of multiple GWAS, which have become much more widely available with much large sample sizes, as the largest one so far for AD with about 74,000 subjects that we will use in our analysis (Lambert et al., 2013).

In summary, we propose a novel gene-based approach to integrating a reference imaging GWAS dataset with clinical GWAS individual-level data or GWAS association summary statistics, differing from existing and popular approaches to treating imaging GWAS data and clinical GWAS data separately. The reference GWAS data can be based on various imaging modalities, as long as the resulting endophenotypes are hypothesized or known to be related to the trait in the main GWAS data. As convincingly demonstrated in TWAS, our proposed IWAS is expected not only to boost statistical power, but also to facilitate biological interpretation of subsequent GWAS discoveries. The proposed tests are developed under a general and rigorous framework of generalized linear models, thus can accommodate various types of quantitative, categorical and survival phenotypes and can adjust for covariates. They are also applicable to both individual-level genotypic, phenotypic data and GWAS summary statistics.

2. Methods

2.1. TWAS

There are two steps in TWAS. Since TWAS is a gene-based method, testing genes one by one, for the purpose of presentation we consider only one gene. The first step is to build a prediction model for the genetic component of the gene's expression level, called "genetically regulated expression (GReX)", by using the possibly *cis*-acting genotypes around the gene based on a reference eQTL dataset. Suppose that in an eQTL dataset, Y^* and $X^* = (X_1^*, \ldots, X_p^*)'$ are the expression level of and the p SNP genotype scores (with additive coding) around the gene. A linear model is assumed: $Y^* = \sum_{j=1}^{p} w_j X_j^* + \varepsilon$, where w_j is the *cis*-effect of SNP j on gene expression and ε is the noise. Based on the eQTL dataset, one can use a method, e.g. elastic net (Zou and Hastie, 2005) as used in PrediXcan, to obtain the estimates $\widehat{w}_j$'s. Next, for a given and separate GWAS dataset, for each subject i with the genotype scores $X_i = (X_{i,1}, \ldots, X_{i,p})'$ around the gene, the predicted GReX is $\widehat{\mathrm{GReX}}_i = \sum_{j=1}^{p} \widehat{w}_j X_{i,j}$. For a trait Y_i for subject i in the GWAS dataset, one simply applies a suitable GLM

$$g(E(Y_i)) = \beta_0 + \widehat{\mathrm{GReX}}_i \beta_c = \beta_0 + \sum_{j=1}^{p} \widehat{w}_j X_{i,j} \beta_c \quad (1)$$

to test for association between the trait and predicted/imputed expression with null hypothesis H_0: $\beta_c = 0$, where $g()$ is the canonical link function (e.g. the logit and the identity functions for binary and quantitative traits respectively), and $E(Y_i)$ is the mean of the trait. Note that for simplicity we do not, but is straightforward to, include covariates in the GLM.

It turns out that TWAS is equivalent to the Sum test (Pan, 2009) in a more general model to test for association between the GWAS trait Y_i and the weighted genotype scores WX_i:

$$g(E(Y_i)) = \beta_0 + \beta' X_i = \beta_0 + \sum_{j=1}^{p} \widehat{w}_j X_{i,j} \beta_j \quad (2)$$

for $H_0: \beta = (\beta_1, \ldots, \beta_p)' = 0$, where $W = \mathrm{Diag}(\widehat{w}_1, \ldots, \widehat{w}_p)$ is a diagonal matrix with the estimated weights as diagonal elements. Xu et al. (2017) showed that, more generally, any more powerful tests, such as an adaptive aSPU test, can be applied. The aSPU test is based on choosing one of the SPU tests, SPU(γ) with positive integers $\gamma \geq 1$ in a candidate set, e.g. $\Gamma = \{1, 2, \ldots, 8, \infty\}$ (Pan et al., 2014). It is known that SPU(1) is equivalent to the Sum test (i.e. PrediXcan), while SPU(2) is a variance-component score test equivalent to kernel machine regression (also known as SKAT in rare variant analysis) with a linear kernel and a nonparametric MANOVA (also called genomic distance-based regression) with the Euclidean distance metric (Wessel and Schork, 2006); the SPU(2) and related tests may yield higher power under many situations (Pan, 2011; Schaid, 2010a, 2010b), as to be confirmed later. In particular, since the Sum test can be derived under the over-simplifying working assumption of $\beta_1 = \beta_2 = \ldots = \beta_p = \beta_c$ in (2) (i.e. all weighted SNPs have an equal effect size and the same effect direction, which is in general incorrect), it clearly shows possible limitations of the Sum and thus PrediXcan (and TWAS): As discussed by others (Pan, 2009; Pan et al., 2014; Wu et al., 2011), the Sum test may lose power if the effect directions of the (weighted) SNPs are different, or the effect sizes are sparse (i.e. with many 0s). More generally, the SPU(γ) tests with $\gamma \in \Gamma = \{1, 2, \ldots, 6, \infty\}$ can be applied, and their results are combined by the aSPU test.

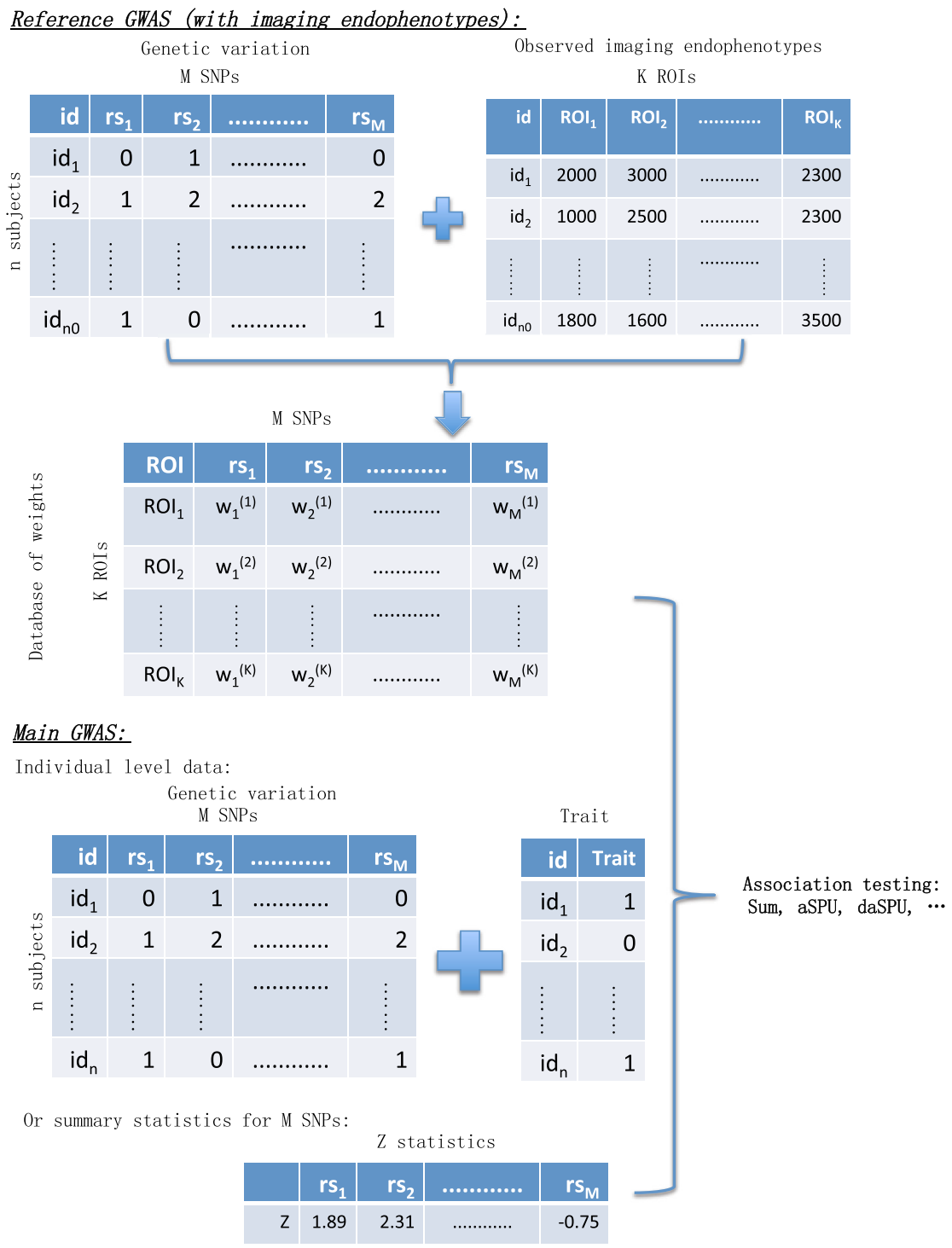

Fig. 1. The workflow of IWAS. It involves two input datasets, a reference GWAS dataset with imaging endophenotypes, and a main GWAS dataset. The reference GWAS dataset contains individual-level data, used to construct a predictive model for one or more imaging endophenotypes based on SNPs, thus giving one or more sets of the weights on SNPs, which are then used along with the main GWAS (either individual-level data or summary statistics) for gene-based association testing.

2.2. IWAS

The main idea of IWAS is similar to TWAS, but instead of using gene expression as the endophenotype to derive the weights, we use an imaging endophenotype; that is, we estimate and then predict/impute the *genetically controlled component* of an imaging endophenotype based on a reference GWAS dataset with imaging endophenotypes. In our examples, we use the gray matter volume of an ROI that is hypothesized to be related to the GWAS trait. Specifically, for example, we have a reference GWAS dataset with both individual-level genotypes and individual-level structural MRI data, from which we construct a prediction model for the imaging endophenotype (e.g. gray matter volume in left hippocampus) using the SNPs around the gene, obtaining the weight matrix W. Then we incorporate the use of W in the main GWAS dataset as for TWAS. The workflow of IWAS is shown in Fig. 1, similar to TWAS (with imaging endophenotypes replaced by gene expression endophenotypes) as shown

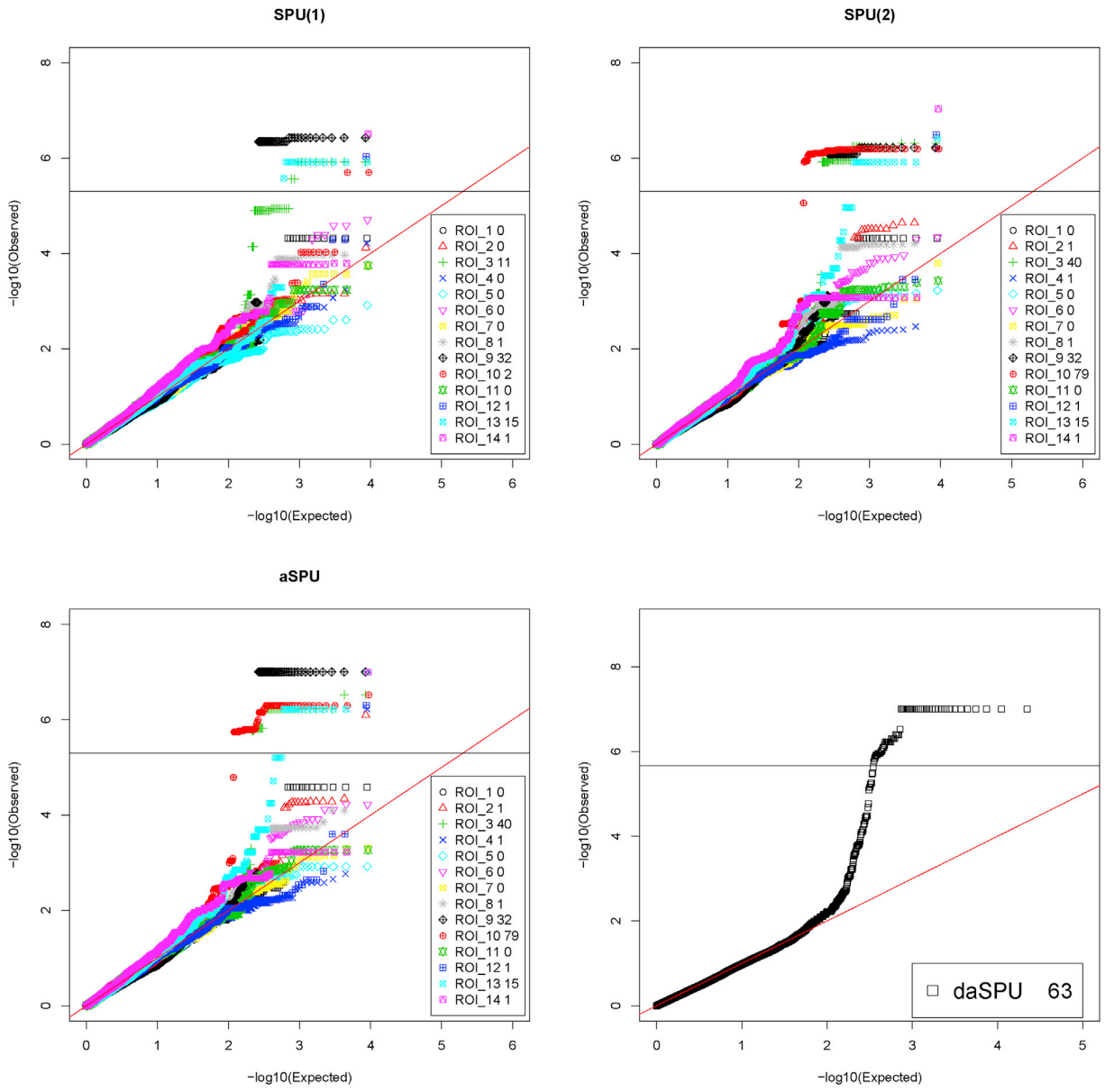

Fig. 2. Q-Q plots for IWAS testing for gene-AD association with the (a) Sum or SPU(1) test, (b) SPU(2) test and (c) aSPU test with weights constructed from gray matter volumes in each of 14 brain ROIs based on the ADNI-1 data, (d) the daSPU test combining the 12 weights from the 12 DMN-related ROIs, applied to the ADNI-GO/2 GWAS data. The numbers in the second column in each legend box of each panel indicate the numbers of the genome-wide significant genes identified by each method.

in Fig. 2 of Gamazon et al. (2015).

2.3. *Association testing with multiple sets of weights/endophenotypes*

There may be compelling reasons to take advantage of multiple sets of weights based on multiple correlated endophenotypes. First, the statistical advantages of joint analysis of multiple traits include possibly increasing statistical power and more precise parameter estimates, alleviating the burden of multiple testing. Biologically, joint analysis of multiple traits addresses the issue of pleiotropy (i.e. one locus influencing multiple traits), providing biological insight into molecular mechanisms underlying the disease or trait. The above conclusions are expected to carry over to the current context of analysis of multiple endophenotypes. Second, a separate application of TWAS or IWAS to each set of weights may lead to inflated type I errors without a suitable multiple testing adjustment, or power loss if the conservative Bonferroni adjustment is used. The omni-bus method of combining multiple sets of weights in TWAS (Gusev et al., 2016) may be difficult to apply due to its need to estimate a large covariance matrix for a large number of the sets of weights. Here we propose extending the aSPU test to the current setting for a large number of the sets of weights, which is expected to be a more common scenario in IWAS. The idea parallels that in other contexts to achieve the data-adaptiveness for multiple purposes (Pan et al., 2015).

Suppose we have K candidate endophenotypes; based on each endophenotype k, we can construct a set of the corresponding weight $\boldsymbol{w}^{(k)} = (w_1^{(k)}, \ldots, w_p^{(k)})'$, $k = 1, 2, \ldots, K$. To avoid the results depending on the varying scales of the sets of weights, we first standardize the weights to have $\sum_{j=1}^{p} \left| w_j^{(k)} \right| = 1$ for each k. Since the performance of any test depends on *unknown* association patterns of the SNPs with the trait (Pan et al., 2014) and on possibly varying (and unknown) informativeness of the endophenotypes/weights, how to use (e.g. by weighting) the SNPs and endophenotypes will largely determine the statistical power of the test. The situation is similar to pathway-based association testing, which critically depends on the unknown association patterns of both the genes (in a pathway) and SNPs (in each gene) (Pan et al., 2015), and to

gene-based association testing on multiple traits, for which one has to account for possibly varying association patterns of the SNPs across the traits (Kim et al., 2016). Accordingly, to maintain high statistical power with various unknown association patterns, as in the above previous works, we use two non-negative integers, γ_1 and γ_2, to control the degrees of weighting over the SNPs and over the endophenotypes respectively, and the two parameters will be chosen data-adaptively as to be shown later. A larger γ_1 puts more weights on the SNPs more likely to be associated with the GWAS trait, while a larger γ_2 upweights the more informative endophenotypes (i.e. those more strongly associated with the GWAS trait). The SPU test statistic is built as follows:

$$\mathrm{SPU}(\gamma_1,\gamma_2)=\sum_{k=1}^{K}\left\{\left[\sum_{j=1}^{p}\left(w_j^{(k)}U_j^*\right)^{\gamma_1}\right]^{\frac{1}{\gamma_1}}\right\}^{\gamma_2},\ U^*=\left(U_1^*,\ldots,U_p^*\right)'$$
$$=\sum_{i=1}^{n}X_i'\left(Y_i-\widehat{\mu}_i^0\right), \tag{3}$$

where U^* is the usual score vector without weighting; we use U^* to emphasize the weighting with weights $w^{(k)}$'s in each SPU test. The parameter γ_1 controls the extent to which to weight each SNP. With $\gamma_1=1$, the SPU test weights each SNP equally, and yields the highest power if all the SNPs (with the given weights from endophenotype k) are associated with the trait with similar effect sizes and the same association direction. When only a subset of SNPs are associated with the trait, or their association directions are different, $\mathrm{SPU}(\gamma_1=2,\gamma_2)$ is often more powerful. As γ_1 increases, $\mathrm{SPU}(\gamma_1,\gamma_2)$ puts heavier weights on the SNPs that are more strongly associated with the trait. At the end, as the parameter approaches to ∞ (as an even integer), it only considers the most significant SNP. Similarly, γ_2 controls how much to up-weight the endophenotypes (i.e. the corresponding sets of weights) that are more informative to the genetic association with the trait of interest. $\mathrm{SPU}(\gamma_1,\gamma_2=1)$ weights all the endophenotypes equally and performs best when each endophenotype is equally informative to the trait. Similarly, as γ_2 increases, the SPU test over-weights the endophenotypes that are more strongly associated with and thus more informative to the trait; in an extreme case, as $\gamma_2\to\infty$, we define $\mathrm{SPU}(\gamma_1,\gamma_2=\infty)=\max_{k=1}^{K}\left\{[\sum_{j=1}^{p}(w_j^{(k)}U_j^*)^{\gamma_1}]^{\frac{1}{\gamma_1}}\right\}$, which takes only the single endophenotype that is most highly associated with the trait. In an extreme case where only one SNP is associated with the trait and only one endophenotype is (dominantly) informative, using $\mathrm{SPU}(\gamma_1=\infty,\gamma_2=\infty)=\max_{j,k}\left|w_j^{(k)}U_j^*\right|$ will be most powerful. Using various combinations of γ_1 and γ_2, one can target and fit different association patterns across multiple SNPs and multiple endophenotypes, yielding higher statistical power.

In practice, because it is unknown which (γ_1,γ_2) value would yield the highest power, we use a **doubly adaptive SPU (daSPU)** test to combine the SPU tests:

$$T_{\mathrm{daSPU}}=\min_{\gamma_1\in\Gamma_1,\gamma_1\in\Gamma_1}P_{\mathrm{SPU}(\gamma_1,\gamma_2)} \tag{4}$$

where $P_{\mathrm{SPU}(\gamma_1,\gamma_2)}$ is the p-value of the $\mathrm{SPU}(\gamma_1,\gamma_2)$ test; empirically we found using $\Gamma_1=\Gamma_2=\{1,2,3,\infty\}$ performed well, which was used in the later analyses.

P-value calculations: A **single** layer/loop of Monte Carlo simulations can be used to obtain the p-values of all the SPU and daSPU tests *simultaneously*. Specifically, we simulate null score vectors $U^{(b)}\sim N(0,V)$ for $b=1,\ldots,B$, from its asymptotic null distribution as multivariate normal with mean 0 and covariance matrix V; there is a closed form solution for V (Pan et al., 2014). The null statistics $\mathrm{SPU}(\gamma_1,\gamma_2)^{(b)}$ can be calculated from the null score vectors $U^{*(b)}$ for $b=1,\ldots,B$. Then the p-value of $\mathrm{SPU}(\gamma_1,\gamma_2)$ is $P_{\mathrm{SPU}(\gamma_1,\gamma_2)}=[1+\sum_{b=1}^{B}(I(\left|\mathrm{SPU}(\gamma_1,\gamma_2)^{(b)}\right|\geq\left|\mathrm{SPU}(\gamma_1,\gamma_2)\right|)/(B+1)]$. And the p-value for the daSPU test can be calculated as $P_{\mathrm{daSPU}}=[1+\sum_{b=1}^{B}I(T_{\mathrm{daSPU}}^{(b)}\leq T_{\mathrm{daSPU}})/(B+1)]$, with $T_{\mathrm{aSPU}}^{(b)}=\min_{\gamma_1\in\Gamma_1,\gamma_1\in\Gamma_1}p_{\gamma_1,\gamma_2}^{(b)}$ and $p_{\gamma_1,\gamma_2}^{(b_1)}=[\sum_{b\neq b_1}I(\left|T_{\mathrm{SPU}(\gamma_1,\gamma_2)}^{(b)}\right|\geq\left|T_{\mathrm{SPU}(\gamma_1,\gamma_2)}^{(b_1)}\right|)+1/(B)]$.

2.4. GWAS summary statistics

Although the above derivations are based the availability of the *main* GWAS individual-level genotypic and phenotypic data, all the tests can be slightly modified and applied to GWAS summary statistics (Kwak and Pan, 2016; Gusev et al., 2016). The main idea is to replace the score vector (based on the individual-level GWAS data) by the Z-scores for the marginal associations of the individual SNPs (from the GWAS summary statistics). However, we do need an individual-level *reference* GWAS dataset with imaging endophenotypes to construct one or multiple sets of weights unless the weights are already given.

3. Results

3.1. ADNI data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by 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, as a 60 million, 5-year public private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

3.2. Imaging and expression endophenotypes

We imputed the ADNI genotype data to the 1000 Genomes Project (1000G) data by the Michigan Imputation Server (Fuchsberger et al., 2015; Howie et al., 2012). Briefly, we imputed the ADNI-1 data (757 subjects, originally based on hg18 and we lifted over to hg19 to facilitate the imputation) and ADNI-GO/2 (793 subjects, originally based on hg19) separately with the following set-up: 1000G Phase 3 v5 as the reference panel, SHAPEIT as the phasing and EUR (European) as the population. The SNPs with minor allele frequency (MAF) > 0.05 and $R^2>0.8$ in both the ADNI-1 and ADNI-GO/2 data were kept. After the imputation and quality control, there were 6,286,276 SNPs. Then all the ambiguous SNPs with alleles A/T or C/G were removed (throughout all analyses in this

paper). Based on the baseline diagnosis, we kept the AD patients and healthy controls (HC) while excluding the subjects with mild cognitive impairment (MCI). The numbers of the subjects in ADNI-1 were 161 and 191 for AD and HC respectively, while those for ADNI-GO/2 were 121 and 150 respectively. For each gene, in addition to its coding region, all variants within the 1Mbp upstream and downstream regions of the transcription start/ending sites were included.

Using the ADNI-1 data as the reference GWAS, we first built a prediction model using elastic net (Zou and Hastie, 2005) as implemented in R glmnet package to obtain a set of weights ($w_j^{(k)}$, $k = 1, \ldots, 14$) for each of the 14 endophenotypes, the gray matter volumes of the 12 regions of interest (ROIs) related to the default mode networks (DMN) (i.e. left and right sides of inferior parietal, inferior temporal, medial orbitofrontal, parahippocampal, precuneus and posterior cingulate) and those of hippocampus, due to the possible relatedness of the above ROIs to AD (Greicius et al., 2004; Saykin et al., 2015). Five covariates were also included: baseline age, gender, baseline education (in years), handedness (left or right), and baseline intracranial volume. In elastic net, as in PrediXcan, we fixed one tuning parameter at 1/2 to give an equal weight to the ridge penalty and Lasso penalty, and the other tuning parameter was selected by five-fold cross-validation with a default setting. To select only those informative genes for each ROI, after selecting the tuning parameters, we calculated the squared Pearson correlation, r^2, between the predicted and observed endophenotype values in the dataset, and selected only those genes with $r^2 > 0.01$; across the 14 sets of the ROIs/weights, the numbers of the selected genes ranged from 9037 to 9997, which were used in later analyses.

We downloaded the tissue-specific gene expression-based weights from the PrediXcan database (https://app.box.com/s/gujt4m6njqjfqqc9tu0oqgtjvtz9860w). Two sets of weights based on GTEx whole blood and brain hippocampus tissues (GTEx, 2015) were used for comparison. For the two tissues, there were 10,210 and 8534 genes with r^2 greater than 0.01, which were tested, respectively.

The Bonferroni adjustment was used to set the genome-wide significance level at $0.05/10000 = 5 \times 10^{-6}$ and $0.05/10500 = 4.76 \times 10^{-6}$ for testing with a single set of weight, as the maximum number of the *mapped* genes for each set of the weights in the following data analyses was less than 10,000 and 10,500 when applied to the ADNI-GO/2 and ADNI-1 cohorts as the main GWAS, respectively; for the daSPU test combining multiple sets of weights, we set the genome-wide significance level at $0.05/23000 = 2.17 \times 10^{-6}$, as the number of *mapped* genes being tested in the analyses was slightly less than 23,000.

We also applied the usual unweighted gene-based tests to the IGAP data. Specifically, we applied the SPU(1), SPU(2) and aSPU tests without weights (i.e. $W = I$) to 23,392 genes; for each gene, all variants within the 20 Kb upstream and downstream regions of the transcription start/ending sites were included. We set the genome-wide significance level at $0.05/23400 = 2.14 \times 10^{-6}$.

We used a step-up procedure to increase the number of simulations, B, when calculating the p-values of the aSPU-type tests in the subsequent data analyses. We started with a small $B = 10^3$, and re-ran the tests with B ten times of its previous value for the genes with a p-value $< 5/B$; we repeated this process up to $B = 10^7$.

3.3. Application to the ADNI data

We first applied IWAS and TWAS to the ADNI-GO/2 data as the main GWAS data by incorporating the weights derived from the ADNI-1 data. Note that, to mimic the realistic situation with separate and independent reference GWAS and main GWAS data, studies, here we used one of the two ADNI datasets to construct the weights, then applied the tests to the other dataset. There were 8917, 8568, 8528, 8874, 9009, 9060, 9134, 8619, 8496, 9397, 9223 and 8712 genes mapped for each set of the weights derived from the 12 ROIs related to DMN, respectively; there were 8990 and 9260 genes mapped for the weights derived from the left and right hippocampus ROIs respectively; there were 8731 and 6944 genes mapped for the weights derived from the whole blood- and hippocampus-specific gene-expression, respectively. After combining the 12 sets of weights for the 12 ROIs related to DMN, we tested a total of 22,282 genes.

As shown in Fig. 2, across the sets of weights (ROIs), the SPU(2) and thus aSPU were more powerful than SPU(1), a test similar to that used in PrediXcan/TWAS. In addition, as expected, using different sets of weights (based on different imaging endophenotypes in the multiple ROIs) yielded quite different numbers of significant genes, demonstrating the necessity and advantage of selecting informative endophenotypes if possible; otherwise, using the proposed method to combine information across multiple sets of weights will be useful. Interestingly, the new daSPU test identified a large number (63) of significant genes.

As a comparison, we applied TWAS with GTEx tissue-specific gene expression as endophenotypes. Five covariates were adjusted: baseline age, gender, education (in years), handedness (left or right), and intracranial volume. Some results for ADNI-GO/2 are shown in Fig. 3; no significant association was detected by any test. It was also confirmed that the SPU(1) test and PrediXcan/TWAS gave essentially the same p-values (not shown).

Finally, we also applied IWAS by reversing the roles of the two ADNI

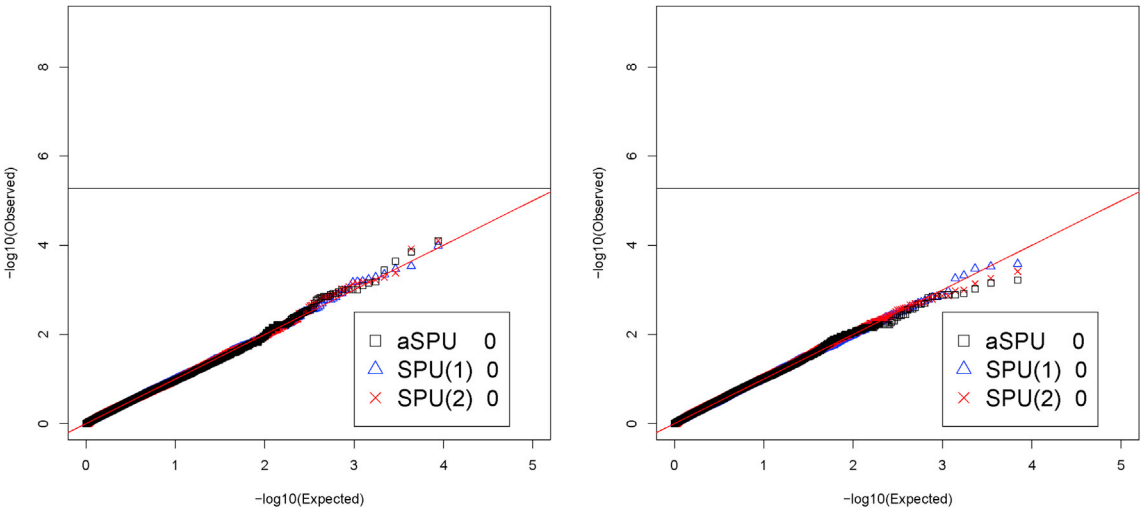

Fig. 3. Q-Q plots for TWAS testing for gene-AD association with weights constructed from the GTEx gene expression in whole blood (left panel) or brain hippocampus (right panel), applied to the ADNI-GO/2 GWAS data. The **numbers** in the second column in each legend box of each panel indicate the *numbers of the genome-wide significant genes* identified by each test.

cohorts: we used the ADNI-GO/2 data as the reference GWAS to construct weights, then applied to the ADNI-1 GWAS data. Across the 14 ROIs, the SPU(1), SPU(2) and aSPU tests detected in total 78, 81 and 44 significant and unique genes respectively, compared to 45, 81 and 63 in the previous analyses; the numbers of the common/intersected sets of the genes from the two analyses were 42, 61 and 37 by the three tests respectively. Furthermore, the daSPU test identified 44 and 63 significant genes with a common set of 37 genes in the two analyses. Hence, in spite of the relatively small sample sizes and some differences between the two ADNI cohorts (Kim et al., 2016), the two analyses yielded largely overlapping sets of the significant genes.

3.4. Application to the IGAP GWAS summary statistics

We applied both IWAS and TWAS to the GWAS summary statistics based on the largest meta-analysis of AD GWAS with over 74,000 individuals (Lambert et al., 2013). The data were downloaded from the International Genomics of Alzheimer's Project (IGAP) site at http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php. We used the IGAP stage 1 summary statistics with a total of 7,055,881 SNPs; the Z-scores were imputed for any missing SNPs using the IMPG algorithm (Pasaniuc et al., 2014). There were 9131, 8901, 8767, 9248, 9540, 9488, 9553, 8954, 8911, 9604, 9695 and 9109 genes mapped for each set of weights derived from the 12 DMN-related ROIs, 8788 and 8818 genes mapped for the left and right hippocampus ROIs, 9308 and 9772 genes mapped for whole blood and hippocampus tissue-specific gene-expression, respectively; after combining the 12 sets of weights for the 12 DMN-related ROIs, we tested on a total of 22,704 genes.

The results are shown in Table 1. First, it is clear that IWAS using the gray matter volumes of several ROIs as endophenotypes yielded much larger numbers of significant genes associated with AD than that of TWAS based on two tissue-specific gene expression. In particular, perhaps due to the relatedness of hippocampus atrophy to AD, IWAS based on the two sides of hippocampus detected the highest numbers of the significant genes. Second, in both IWAS and TWAS, for almost every set of the weights, the adaptive aSPU test identified a larger number of the significant genes than that of SPU(1) (i.e. the Sum test), a similar test used in the original PrediXcan/TWAS. Third, by combining the weights derived from the 12 DMN-related ROIs, the daSPU test yielded a large number (44) of significant genes. Note that, for the purpose of illustration we did not adjust for multiple testing across multiple sets of weights when deciding the significance level for each set of weights. Strictly speaking, such an adjustment is necessary in practice, which may be conservative if a simple Bonferroni adjustment is used. Alternatively, using the daSPU test across multiple sets of weights avoids such a multiple testing problem. Given the relevance of hippocampal atrophy to AD, we did not combine the weights from hippocampus with the 12 DMN-related ROIs. As shown in the Manhattan plots in Fig. 4, the significant genes or loci identified by the daSPU test based on the 12 DMN-related ROIs and those of aSPU on the two hippocampus ROIs are complementary. Finally, among the significant genes identified by IWAS or TWAS, some are novel ones as shown in Tables 2 and 3, which of course are only putative and need to be validated; some of the novel genes are known to be associated with other traits Supplementary table).

It is noted that the usual unweighted aSPU test also detected 50 significant genes, among which only 11 overlaps with the 44 genes identified by the weighted daSPU test, illustrating possible values and a complementary role of IWAS compared to the standard (unweighted) GWAS analysis. It was also confirmed that the unweighted SPU(1) was likely less powerful in identifying fewer significant genes than the unweighted SPU(2) and aSPU tests.

3.5. Application to the lipid GWAS summary data

Since the ROIs were chosen based on their possible relatedness to AD, the corresponding neuro-imaging endophenotypes were expected to

Table 1
The numbers of the significant AD-associated genes identified by IWAS and TWAS with the IGAP data. The numbers a/b in each cell indicate the total number of significant genes/number of significant genes that overlap at least a genome-wide significant SNP.

	IWAS													TWAS		Unweighted
	Inferior parietal		Inferior temporal		Mediala orbitofrontal		Parahippocampal		Precuneus		Posterior cingulate		Hippocampus	GTEx		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Blood	Hippo w/o weights	
SPU(1)	8/8	65/64	41/40	73/64	29/3	51/32	59/47	68/56	26/26	83/79	33/32	71/70	54/54	71/68	13/5	11/4	22/20
SPU(2)	10/10	66/66	40/40	67/67	0/0	56/56	89/81	61/61	44/44	83/81	42/42	70/70	83/83	78/78	15/11	7/5	45/40
aSPU	10/10	70/69	40/40	77/69	7/0	75/56	81/69	76/64	42/42	85/81	42/41	74/73	83/83	77/76	16/9	13/5	50/47
daSPU	44/44																

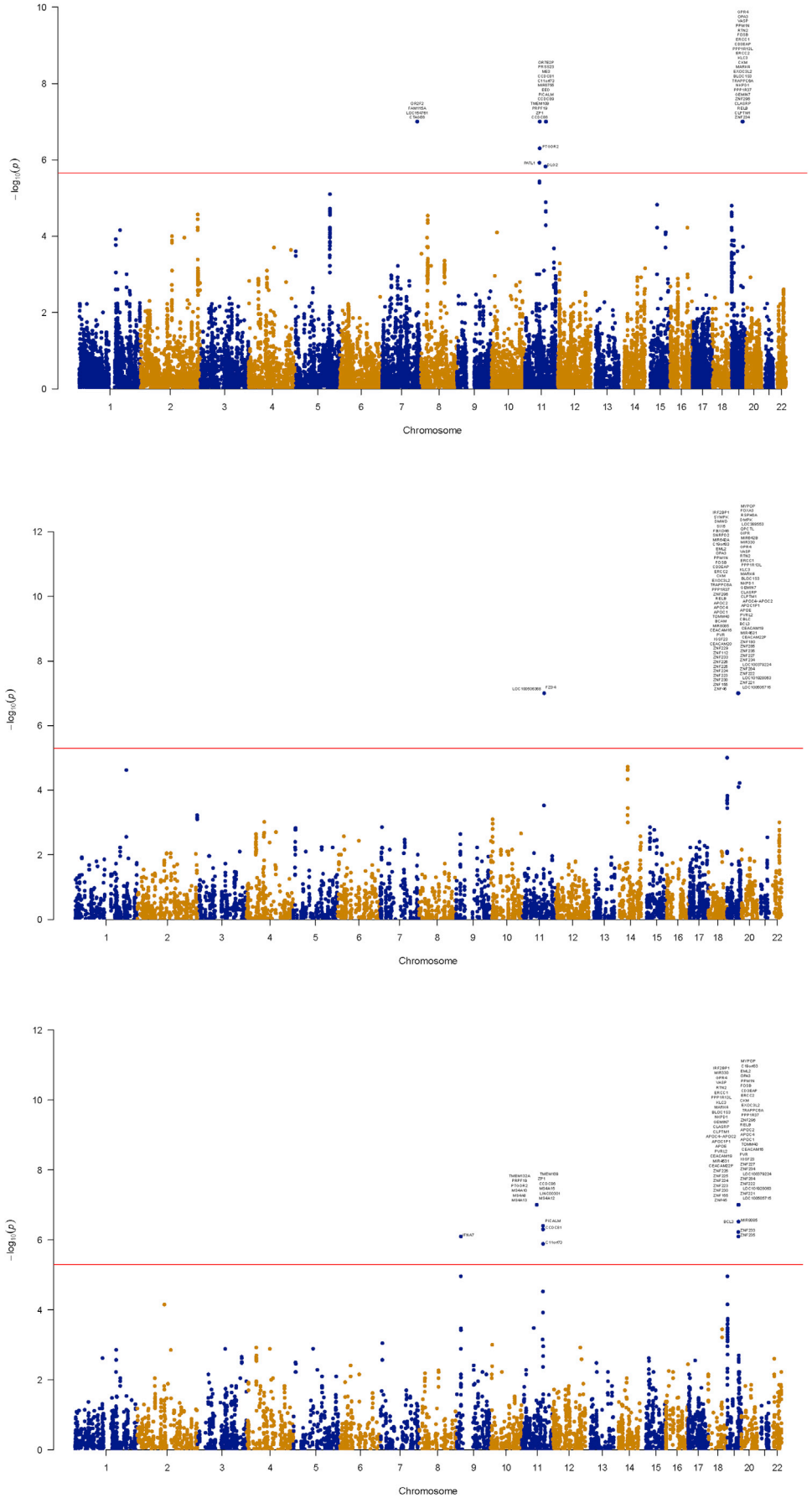

Fig. 4. Manhattan plots of IWAS for the IGAP data: from the top to the bottom, the daSPU based on combining the 12 DMN-related ROIs, aSPU based on the left hippocampus and aSPU on the right hippocampus. The horizontal line in each panel gives the genome-wide significance level.

Table 2
Novel significant AD-associated genes identified by IWAS, not overlapping with any genome-wide significant SNPs within ±1 Mb region based on the IGAP data. Column "#SNPs" gives the numbers of SNPs used in the genes, column "sig.test" indicates which of SPU(1) and aSPU tests detected the genes, while the remaining three columns give the p-values of the three tests.

ROI	chr	genes	#SNPs	sig.test	aSPU	SPU(1)	SPU(2)
10	1	COLGALT2	10	both	1.0E-07	0.0E+00	2.7E-01
10	1	C1orf21	16	both	5.0E-07	1.5E-07	2.8E-01
10	1	PTPRC	3	both	1.0E-07	0.0E+00	1.6E-01
6	2	PRKCE	4	both	1.1E-06	4.8E-07	2.5E-01
4	3	EXOSC7	2	both	1.0E-07	4.6E-12	4.3E-03
4	3	TMEM158	11	both	1.0E-07	0.0E+00	4.5E-03
4	3	LARS2-AS1	16	both	7.0E-07	2.1E-07	5.5E-03
5	3	EPHA6	3	both	1.0E-07	0.0E+00	3.1E-01
5	3	CRYBG3, MINA, GABRR3, OR5AC2, OR5H1 OR5H14, OR5H15, OR5H6, OR5H2, OR5K4, OR5K1, OR5K2, CLDND1, GPR15, CPOX ST3GAL6-AS1, ST3GAL6, DCBLD2, OR5K3	2	SPU(1)	6.0E-06	3.8E-06	3.0E-02
2	4	GALNT7	2	both	1.0E-07	9.3E-09	4.9E-01
10	4	ADAM29	2	both	1.0E-07	0.0E+00	2.9E-01
7	5	LINC01024, PURA, IGIP, CYSTM1	5	both	4.6E-06	4.0E-06	9.5E-06
7	5	WDR55, DND1, HARS, HARS2, ZMAT2 VTRNA1-1, VTRNA1-2, VTRNA1-3	7	both	9.0E-07	3.3E-07	1.1E-06
8	8	CLVS1	2	both	1.0E-07	8.9E-14	1.5E-02
14	9	IFNA7	41	both	8.0E-07	5.0E-07	1.3E-02
14	9	IFNA10, IFNA16	41	SPU(1)	1.1E-05	3.1E-06	1.4E-02
6	9	ANKRD19P, ZNF484,LOC642943,FGD3, C9orf89, NINJ1, C9orf129, FAM120AOS, PHF2, MIRLET7A1, MIRLET7F1, WNK2, MIRLET7D, LOC100132077, HIATL1, SUSD3, MIRLET7DHG, FAM120A	2–5	both	1.0E-07	0.0E+00	1.3E-01
4	9	PSMD5-AS1	3	both	1.0E-07	4.8E-12	3.1E-01
4	10	DNAJC1	2	SPU(1)	1.7E-05	2.7E-06	3.1E-02
11	10	AGAP7	5	both	1.1E-06	5.4E-07	1.4E-02
5	16	GRIN2A	3	both	1.0E-07	0.0E+00	4.2E-02
5	16	ATF7IP2	4	both	1.0E-07	7.3E-10	4.2E-02
8	16	ARHGAP17, LOC554206, LCMT1, LOC100506655, AQP8, ZKSCAN2 LOC102723510	2	both	1.0E-07	0.0E+00	7.3E-01
12	16	DDX19A	5	both	3.5E-06	1.8E-06	2.2E-05
5	17	MAP2K6, LOC101928122	3	both	1.0E-07	0.0E+00	3.6E-01
5	17	LOC101928122	3	both	1.0E-07	0.0E+00	3.6E-01
5	17	LOC102723487,LINC01028	3	both	1.0E-07	9.3E-13	3.7E-01
3	18	DSEL	3	SPU(1)	5.4E-06	2.5E-06	6.5E-01
8	19	AP1M1, MED26, SMIM7, TMEM38A	2	both	1.0E-07	0.0E+00	7.2E-02

enhance statistical power in detecting larger numbers of significant AD-associated genes, as shown in the above example. However, for an unrelated GWAS trait, the above neuro-imaging endophenotypes would not be expected to improve the power. To confirm, we applied IWAS (with ROI #9) and TWAS (with the GTEx whole blood gene-expression) to a meta-analyzed large lipid GWAS summary statistics dataset for trait HDL (Global Lipids Genetics Consortium, 2013); the Z-scores were imputed for any missing SNPs present in the reference GWAS using the IMPG algorithm (Pasaniuc et al., 2014). There were 9455 and 8916 genes mapped for the weights derived from GTEx whole blood gene-expression and ROI # 9, respectively.

As expected, the number of the significant genes identified by IWAS was no more than (in fact, less than) that by TWAS (Fig. 5). Note that, due to the large sample size (about 189,000 individuals) for the lipid data, significant associations are expected with even a non-informative set of weights.

4. Discussion

We have developed an IWAS approach to integrating imaging endophenotypes with GWAS to identify genes associated with a complex trait or disease. Using the ADNI-1 data as a reference GWAS dataset with

Table 3
Novel significant AD-associated genes identified by TWAS, not overlapping with any genome-wide significant SNPs within ±1 Mb region based on the IGAP data.

Tissue	chr	gene	start	end	#SNPs	sig.test	aSPU	SPU(1)	SPU(2)
hippo	14	ENTPD5	73433180	75486026	2	aSPU	2.9E-06	8.5E-01	4.4E-07
wb	7	SLC26A3	106405911	108443678	8	SPU(1)	6.0E-06	2.5E-06	5.9E-02
wb	17	CHRNE	3801063	5806369	10	SPU(1)	7.0E-06	2.7E-06	2.2E-06
hippo	1	DENND4B	152901976	154919154	10	both	4.0E-07	4.6E-08	4.9E-03
wb	5	SLC4A9	138739786	140754722	1	both	2.0E-07	1.1E-07	1.1E-07
wb	7	TAS2R4	140478288	142479188	4	both	1.0E-07	0.0E+00	5.5E-02
wb	9	CDC37L1	3679565	5706594	4	both	1.0E-07	1.7E-13	6.7E-02
hippo	10	SEMA4G	101732285	103745373	2	both	1.0E-07	0.0E+00	1.8E-01
hippo	11	IFITM10	753639	2771824	2	both	1.0E-07	1.4E-08	3.0E-01
hippo	11	TRAPPC4	117889240	119894385	10	both	5.0E-07	2.7E-07	3.1E-01
hippo	15	LDHAL6B	58499014	60500785	2	both	1.0E-07	0.0E+00	6.4E-01
wb	16	DDX19A	69380823	71407281	2	both	1.0E-07	0.0E+00	5.6E-01
wb	19	GNA11	2094407	4124000	5	both	1.8E-06	5.5E-07	3.6E-01
hippo	19	SSBP4	17530145	19545372	7	both	3.2E-06	1.7E-06	4.5E-06
wb	20	TRIB3	0	1378203	15	both	1.0E-06	3.9E-07	1.1E-02
hippo	20	YWHAB	42514239	44537175	2	both	1.0E-07	0.0E+00	6.6E-01

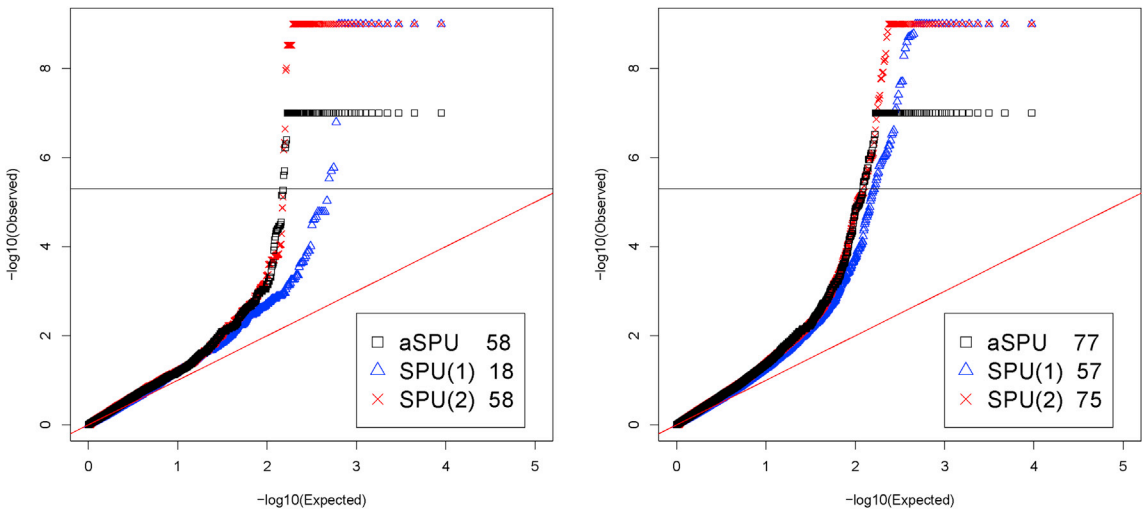

Fig. 5. Q-Q plots for (a) IWAS and (b) TWAS testing for gene-HDL association applied to the 2013 lipid GWAS summary statistics. The **numbers** in the second column in each legend box of each panel indicate the *numbers of the genome-wide significant genes* identified by each method.

individual-level genotypic data and structural MRI-derived gray matter volumes of some ROIs possibly related to AD, we applied IWAS to the ADNI-GO/2 GWAS individual-level data and a GWAS summary statistics dataset, uncovering multiple genes, including some novel ones, significantly associated with AD. We also compared its performance with TWAS with gene expression as endophenotypes. Perhaps due to more direct relatedness of brain atrophy, instead of gene expression in various tissues of normal subjects, to AD, our IWAS consistently identified larger numbers of AD-associated genes than TWAS.

As in TWAS, the choice of an endophenotype to construct a set of weights to be used in IWAS is important. Depending on the choice, IWAS may or may not increase the statistical power compared to the standard GWAS without using the endophenotype. In both IWAS and TWAS, only when an endophenotype is intermediate between the genetics and the trait, or when the genetic component of the trait has a causal effect on the endophenotype, integrating the endophenotype into the main GWAS may boost the statistical power to detect the trait-associated genes. In our example, some ROIs like hippocampus led to much larger numbers of significant genes than did other ROIs in DMN; even though DMN is known to be likely associated with AD, some ROIs may be more informative (e.g. more strongly associated with AD and with a high heritability) than other ROIs in DMN. Hence, in practice one has to be careful in selecting the endophenotypes to be used, e.g. based on the endophenotype ranking value (ERV) proposed by Glahn et al. (2012). In realistic situations, one can do his/her best to select all relevant endophenotypes, then check the results of IWAS for each endophenotype with a suitable control of multiple testing, e.g. by a conservative Bonferroni adjustment. Alternatively, one can also apply our proposed daSPU test to combine the results across multiple endophenotypes. Finally, instead of using structural MRI-derived imaging endophenotypes, one can use other imaging modalities to derive endophenotypes as long as it is reasonable to assume that they are possibly related to the trait of interest. For example, it may be worthwhile to apply IWAS with brain functional or structural connectivities (Thompsona et al., 2013) as endophenotypes.

The proposed statistical tests used in IWAS and TWAS are implemented in R package aSPU2 that will be publicly available on CRAN; the sets of the weights derived from the ADNI-1 imaging endophenotypes and example computer code are publicly available at wuchong.org/IWAS.html.

Acknowledgment

The authors are grateful to the reviewers for helpful and constructive comments. This research was supported by NIH grants R21AG057038, R01HL116720, R01GM113250 and R01HL105397, and by the Minnesota Supercomputing Institute; ZX was supported by a University of Minnesota MnDRIVE Fellowship and CW by a University of Minnesota Dissertation Fellowship.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Rev December 5, 2013 Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (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 Southern California.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.neuroimage.2017.07.036.

References

Balthazar, M., Weiler, M., Campos, B., Rezende, T., Damasceno, B., Cendes, F., 2014. Alzheimer as a Default Mode Network Disease: a grey matter, functional and structural connectivity study. Neurology 83 (10). P6.324.

Bertram, L., Tanzi, R.E., 2011. Genetics of Alzheimer's disease. In: Dickson, D.W., Weller, R.O. (Eds.), Neurodegeneration: the Molecular Pathology of Dementia and Movement Disorders, second ed. Wiley-Blackwell, Oxford, UK. http://dx.doi.org/10.1002/9781444341256.ch9.

Damoiseaux, J.S., Seeley, W.W., Zhou, J., Shirer, W.R., Coppola, G., Karydas, A., Rosen, H.J., Miller, B.L., Kramer, J.H., Greicius, M.D., Alzheimer's disease neuroimaging initiative, 2012. Gender modulates the APOE ε4 effect in healthy older

adults: convergent evidence from functional brain connectivity and spinal fluid tau levels. J. Neurosci. 32, 8254–8262.
Fuchsberger, C., Abecasis, G.R., Hinds, D.A., 2015. minimac2: faster genotype imputation. Bioinformatics 31, 782–784.
Gamazon, E.R., et al., 2015. A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. 47, 1091–1098.
1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073.
Glahn, D.C., Winkler, A.M., Kochunov, P., Almasy, L., Duggirala, R., Carless, M.A., Curran, J.C., Olvera, R.L., Laird, A.R., Smith, S.M., et al., 2010. Genetic control over the resting brain. Proc. Natl. Acad. Sci. 107 (3), 1223–1228.
Glahn, D.C., Curran, J.E., Winkler, A.M., Carless, M.A., Kent Jr., J.W., Charlesworth, J.C., Johnson, M.P., Gring, H.H., Cole, S.A., Dyer, T.D., Moses, E.K., Olvera, R.L., Kochunov, P., Duggirala, R., Fox, P.T., Almasy, L., Blangero, J., 2012. High dimensional endophenotype ranking in the search for major depression risk genes. Biol. Psychiatry 71, 6–14.
Global Lipids Genetics Consortium, 2013. Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283.
Gottesman II, Gould, T.D., 2003. The endophenotype concept in psychiatry: etymology and strategic intentions. Am. J. Psychiatry 160, 636–645.
Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc. Natl. Acad. Sci. 101, 4637–4642.
GTEx Consortium, 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660.
Gusev, A., et al., 2016. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. 48, 245–252.
He, Y., Chen, Z., Gong, G.L., Evans, A., 2009. Neuronal networks in Alzheimer's disease. Neuroscientist 15, 333–350.
Hong, M.G., Reynolds, C.A., Feldman, A.L., Kallin, M., Lambert, J.C., Amouyel, P., Ingelsson, E., Pedersen, N.L., Prince, J.A., 2012. Genome-wide and gene-based association implicates FRMD6 in Alzheimer disease. Hum. Mutat. 33, 521–529.
Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., Abecasis, G.R., 2012. Fast and accurate genotype imputation in genome-wide association studies through prephasing. Nat. Genet. 44, 955–959.
Huang, M., Nichols, T., Huang, C., Yu, Y., Lu, Z., Knickmeyer, R.C., Feng, Q., Zhu, H., Alzheimer's Disease Neuroimaging Initiative, 2015. FVGWAS: fast voxelwise genome wide association analysis of large-scale imaging genetic data. Neuroimage 118, 613–627.
Jones, D.T., Machulda, M.M., Vemuri, P., McDade, E.M., Zeng, G., Senjem, M.L., Gunter, J.L., Przybelski, S.A., Avula, R.T., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack Jr., C.R., 2011. Age-related changes in the default mode network are more advanced in Alzheimer disease. Neurology 77 (16), 1524–1531.
Karch, C.M., Cruchaga, C., Goate, A.M., 2014. Alzheimer's disease genetics: from the bench to the clinic. Neuron 83 (1), 11–26.
Kim, J., Zhang, Y., Pan, W., 2016. Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data. Genetics 203 (2), 715–731.
Kwak, I., Pan, W., 2016. Adaptive gene- and pathway-trait association testing with gwas summary statistics. Bioinformatics 32, 1178–1184.
Lambert, J.C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M.J., Tavernier, B., et al., 2009. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat. Genet. 41 (10), 1094–1099.
Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. 45, 1452–1460.
Lin, D., Cao, H., Calhoun, V.D., Wang, Y.P., 2014. Sparse models for correlative and integrative analysis of imaging and genetic data. J. Neurosci. Methods 237, 69–78.
Lu, Z.-H., Khondker, Z., Ibrahim, J.G., Wang, Y., Zhu, H., for the Alzheimer's Disease Neuroimaging Initiative, 2017. Bayesian longitudinal low-rank regression models for imaging genetic data from longitudinal studies. NeuroImage 149, 305–322.
Marei, H., Althani, A., El Zowalaty, M., Albanna, M.A., Cenciarelli, C., Wang, T., Caceci, T., 2016. Common and rare variants associated with Alzheimer's disease. J. Cell Physiol. 231 (7), 1432–1437.
Metin, B., Krebs, R.M., Wiersema, J.R., Verguts, T., Gasthuys, R., van der Meere, J.J., Achten, E., Roeyers, H., Sonuga-Barke, E., 2015. Dysfunctional modulation of default mode network activity in attention-deficit/hyperactivity disorder. J. Abnorm Psychol. 124 (1), 208–214.
Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., Cox, N.J., 2010. Trait-associated SNPs are more likely to Be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 6, e1000888.
Pan, W., 2009. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. Genet. Epidemiol. 33, 497–507.
Pan, W., 2011. Relationship between genomic distance-based regression and kernel machine regression for multi-marker association testing. Genet. Epidemiol. 35 (4), 211–216.
Pan, W., Kim, J., Zhang, Y., Shen, X., Wei, P., 2014. A powerful and adaptive association test for rare variants. Genetics 197, 1081–1095.
Pan, W., Kwak, I.-Y., Wei, P., 2015. A powerful pathway-based adaptive test for genetic association with common or rare variants. Am. J. Hum. Genet. 97, 86–98.
Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., Hirschhorn, J., Strachan, D.P., Patterson, N., Price, A.L., 2014. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. Bioinformatics 30, 2906–2914.
Ridge, P.G., Mukherjee, S., Crane, P.K., Kauwe, J.S., Alzheimer's Disease Genetics Consortium, 2013. Alzheimer's disease: analyzing the missing heritability. PLoS ONE 8, e79771.
Saykin, A.J., Shen, L., Yao, X., Kim, S., Nho, K., Risacher, S.L., Alzheimer's Disease Neuroimaging Initiative, 2015. Genetic studies of quantitative MCI and AD phenotypes in ADNI: progress, opportunities, and plans. Alzheimer's Dementia 11, 792–814.
Schaid, D.J., 2010a. Genomic similarity and kernel methods I: advancements by building on mathematical and statistical foundations. Hum. Hered. 70, 109–131.
Schaid, D.J., 2010b. Genomic similarity and kernel methods II: methods for genomic information. Hum. Hered. 70, 132–140.
Shen, L., Thompson, P.M., Potkin, S.G., Bertram, L., Farrer, L.A., Foroud, T.M., for the Alzheimer's Disease Neuroimaging Initiative, 2014. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. Brain Imaging Behav. 8, 183–207.
Sherva, R., Tripodis, Y., Bennett, D.A., Chibnik, L.B., Crane, P.K., de Jager, P.L., Farrer, L.A., Saykin, A.J., Shulman, J.M., Naj, A., , et al.GENAROAD Consortium, Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Genetics Consortium, 2014. Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. Alzheimer's Dement. 10, 45–52.
Tao, C., Nichols, T.E., Hua, X., Ching, C.R.K., Rolls, E.T., Thompson, P.M., Feng, J., The Alzheimer's Disease Neuroimaging Initiative, 2017. Generalized reduced rank latent factor regression for high dimensional tensor fields, and neuroimaging-genetic applications. NeuroImage 144, 35–57.
Thompsona, P.M., Ge, T., Glahn, D.C., Jahanshada, N., Nichols, T.E., 2013. Genetics of the connectome. NeuroIamge 80, 475–488.
Wessel, J., Schork, N.J., 2006. Generalized genomic distance-based regression methodology for multilocus association analysis. Am. J. Hum. Genet. 79, 792–806.
Wu, M.C., Lee, S., Cai, T., et al., 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. 89, 82–93.
Xu, Z., Wu, C., Wei, P., Pan, W., 2017. A Powerful Framework for Integrating eQTL and GWAS Summary Data.
Zhu, H., Khondker, Z., Lu, Z., Ibrahim, J.G., 2014. Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. J. Am. Stat. Assoc. 109, 977–990.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B 67, 301–320.