Statistical Methods for Imaging Genetics

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Junghi Kim

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Advised by Wei Pan, Ph.D

May, 2016

Acknowledgements

I am grateful to my advisor, Professor Wei Pan, for his guidance and encouragement.

Abstract

Neuroimaging phenotypes are often collected in genome-wide association studies (GWASs) as secondary phenotypes for a disease outcome. Joint analysis of multivariate imaging phenotypes can incorporate neural activity from multiple brain regions, and boost statistical power in association analysis by taking advantage of similarity across phenotypes. Yet, most GWASs are based on case-control study designs, implying that regression approaches not adjusted for the sampling design may lead to biased estimates of associations for secondary phenotypes with inflated Type I error rates and reduced power. Despite this well-known result, unadjusted regression models are widely used in the imaging genetic literature. The aim of this thesis is twofold: 1) to identify the conditions when sampling bias occurs in association analysis of secondary phenotypes, and 2) to improve power for gene discovery, utilizing multiple imaging phenotypes.

Potential bias introduced by the unadjusted regression model is demonstrated using the Alzheimer's Disease Neuroimaging Initiative (ADNI) data. In simulation studies, we compare the performance of the naive approach with those of several existing methods accounting for ascertainment bias to demonstrate potential issues in using standard analyses of secondary phenotypes. Finally we propose two novel statistical methods to identify genetic associations with multiple phenotypes to improve testing power. The first method is to detect single-SNP and multi-trait associations in a proportional odds model (POM). The second considers multi-SNP and multi-trait associations in the generalized estimating equations (GEE) framework, applied to rare variants in sequencing data and pathway analysis. Both methods extend the recently proposed adaptive sum of powered score (aSPU) test, shown to maintain high power in a wide range of situations. New methods are demonstrated in real data analyses and simulation studies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Imaging genetics leverages the strengths of both neuroimaging and genetic studies [1, 2]. While the analysis of association between a primary outcome (disease status) and genetic factors is generally the main focus of the study, in imaging genetic studies hundreds to thousands of neuroimaging and neuropsychological phenotypes are collected as secondary phenotypes. Alzheimer's Disease Neuroimaging Initiative (ADNI) was started in 2004 and is being continued since, collecting extensive clinical, genomic and multi-modal imaging data to advance our understanding of the initiation, progression and etiology of Alzheimer's disease (AD) [2]. With the availability of the secondary phenotypes, there is an increasing interest in developing statistical methods to incorporate synchronous brain activities in multiple brain regions by employing multiple imaging phenotypes. Multiple phenotypes often measure the same underlying trait; by taking advantage of similarity across phenotypes, one could potentially boost statistical power in association analysis. In imaging genetic studies often a secondary phenotype is treated as an intermediate phenotype for the disease. Hence the use of secondary phenotypes provides some advantages over that of a disease status, both in improving power for discovering risk genes and in understanding underlying pathogenic mechanisms of neurological disorder like AD. For example, APOEε4 allele has been consistently shown to be associated with AD. However, only 50% of AD patients carry an APOEε4 allele, suggesting the existence of other genetic variants contributing to risk for the disease [3]; a recent study

indicates that more than 25% of phenotypic variance remains unexplained by known genetic factors [4], implying many more genes underlying late onset AD are waiting to be discovered.

This thesis focuses on genetic associations for multiple imaging phenotypes: embracing issues arising from current research using secondary phenotypes and developing two novel statistical methods using multiple imaging phenotypes. For the first work, we illustrate a cautionary note on using secondary phenotypes in imaging genetics. Analyses with the secondary phenotypes are not straightforward, because including ADNI, most genome-wide association studies (GWASs) with secondary phenotypes are based on the case-control study design, implying that the resulting case-control data are likely a biased, not random, sample of the target population. A major characteristic of a case-control study is that the disease-status is identified at the beginning of the study, and the sampling of the subjects is conditional on their disease status. For instance, ADNI collected its samples based on participants' disease status: specifically, 200 healthy controls (HCs), 400 subjects with mild cognitive impairment (MCI) and 200 patients with AD were recruited; the set of the ADNI participants is not expected to be a random sample of the age-matched general population. For instance, in the general population, ten to twenty percent of people age 65 or older is known to have MCI [5, 6, 7], but nearly a half of the ADNI samples consists of MCI individuals. Hence, although a standard logistic regression model can be applied to draw unbiased inference for genetic associations with the risk of AD, a standard regression model (without any suitable adjustment) may lead to biased inference of genetic associations. Despite of this well known result in genetic epidemiology, to our surprise, all the published studies on secondary phenotypes with the ADNI samples have ignored this potential problem [8, 9, 10, 11]. We aim to answer whether such a standard analysis of a secondary phenotype is valid or problematic using the ADNI data as an example. A number of studies have been proposed for valid analysis of secondary phenotypes, accounting for biased case-control sampling [12, 13, 14]. By comparing with the valid methods in both real data analyses and simulation studies, we found that such a standard regression model was generally valid (with only small biases or slightly inflated Type I errors) for the ADNI data, though cautions must be taken when analyzing other data. The main reason for our conclusion to hold for the ADNI data (and possibly other data) is due to the high prevalence of the AD (or other disease)

in the target population, leading to its small difference from the sampling proportion of the cases in the case-control sample.

Bearing in mind that a standard regression model is generally valid for a secondary phenotype with ADNI samples, we propose two novel statistical methods to detect genetic associations with multiple imaging phenotypes. The first method is for single-SNP and multi-trait associations; the second considers SNP set-based associations deciphering complicated joint effects of multiple SNPs on multiple traits, which can be applied to the gene- and pathway-based analyses. A fundamental challenge in multi-trait analysis is the lack of a uniformly most powerful test. A key issue is how to maximize the statistical power in the presence of many non-associated traits, while gaining the power when many or most of the traits are weakly associated. In the former situation, one can avoid losing testing power by utilizing only few top associated traits [15, 16, 17] or dimension reduction methods [18, 19, 20, 21, 22, 23, 24]. In contrast, in the latter situation with many weak associations, burden tests [8, 25] or variance component tests [26, 27] are preferred, in which multiple traits are jointly analyzed. Yet the true association pattern is unknown in practice, and a statistical method has to be flexible enough to adapt to the given data to consider the two extreme situations. One example is an adaptive test for single SNP-multi trait associations introduced by Zhang et al. [28]: the adaptive test is able to capture joint associations of multiple traits with dense association signals while maintaining high statistical power even with sparse association patterns, by upweighting the traits more highly associated with the SNP and vice versa. The two proposed tests are built on such an adaptive test to take advantage of its benefits.

First, we present a novel statistical method for single SNP-multi trait associations. Including Zhang et al. [28], most of the existing methods treat multiple traits as responses while treating an SNP as a predictor coded under an additive inheritance mode [12, 15, 20, 28]. Instead, we regard an SNP as an ordinal response while treating traits as predictors in a proportional odds model (POM). In this way, it is not only easier to handle mixed types of traits, e.g. some quantitative and some binary, but also potentially more robust to the commonly adopted additive inheritance mode. More importantly, we develop an adaptive test in a POM so that it can maintain high power across many possible situations. Compared to the existing methods treating multiple traits as responses, e.g. in a generalized estimating equation (GEE) approach [28], the

proposed method can be applied to a high dimensional setting where the number of phenotypes (p) can be larger than the sample size (n), in addition to a usual small p setting. The promising performance of the proposed method was demonstrated with applications to the ADNI data, in which either structural MRI driven phenotypes or resting-state functional MRI (rs-fMRI) derived brain functional connectivity were used as phenotypes. The applications led to the identification of several top SNPs of biological interest. Furthermore, a simulation study showed competitive performance of the new method compared to several existing methods, including potential power gain of the new method in cases with a dominant inheritance mode.

Finally, we propose a novel gene- and pathway-based association test for multiple phenotypes, which extends the adaptive testing under the GEE framework [28]. Due to the well-known genetic heterogeneity and small effect sizes of individual SNPs, as observed from published GWAS results [29], it may be promising to conduct association analysis at the SNP-set (or gene) or pathway level, rather than at the individual SNP level, in identifying aggregate effects of multiple SNPs. Based on the general framework of GEE, Zhang et al. [28] is flexible in incorporating covariates and various types of traits [30], and able to avoid correctly specifying a joint multivariate distribution or likelihood for a set of multiple traits. The proposed method has the same spirit as Zhang et al. [28] for using GEE. Because the power of a test critically depends on several unknown factors such as the proportions of associated SNPs and of traits, the proposed test is adaptive at both the SNP and trait levels, giving larger weights to those likely associated SNPs and traits to yield high power across a wide spectrum of situations. We illuminate on relationships among the proposed and some existing tests, showing that the proposed test covers several existing tests as special cases. We compare the performance of the new test with several existing tests using both simulated and real data. The methods were applied to structural MRI data drawn from ADNI to identify genes associated with grey matter atrophy in the human brain default mode network (DMN). We applied the method to rare variants in sequencing data and extended to pathway analysis, using adaptive weighting at the gene-level, in addition to at the SNP- and trait-levels.

Chapter 2 addresses whether standard linear regression of secondary phenotypes would lead to biased estimates of association for secondary phenotypes with inflated

Type I errors and reduced power [31]. Chapter 3 and Chapter 4 introduce new statistical methods for genetic associations with multiple imaging phenotypes: Chapter 3 discusses a novel test for single SNP-multi trait associations in POM, while Chapter 4 proposes the novel test of multi-trait and multi-SNP associations for GWAS and sequencing data [32].

1.2 Data

Data used in the preparation of this thesis 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`.

1.3 A list of publications

The following list of publications is related to my Ph.D studies and research.

Kim, J., Zhang, Y. and Pan, W. (2016) Powerful and adaptive testing for multitrait and multi-SNP associations with GWAS and sequencing data. To appear in *Genetics*.

Kim, J. and Pan, W. (2015) Highly adaptive tests for group differences in brain connectivity. *NeuroImage: Clinical* 9:625-639.

Kim, J., Bai, Y. and Pan, W. (2015) An adaptive association test for multiple phenotypes with GWAS summary statistics. *Genetic Epidemiology* 39 (8):651-663. co-first authors

Kim, J. and Pan, W. (2015) A cautionary note on using secondary phenotypes in neuroimaging genetic studies. *NeuroImage* 121:136-145.

Kim, J., Wozniak, JR., Mueller BA., and Pan, W. (2015). Testing group differences in brain functional connectivity: using correlations or partial correlations? *Brain Connectivity* 5 (4):214-231.

Kim, J., Wozniak, JR., Mueller, BA., Shen, X., and Pan, W. (2014) Comparison of statistical tests for group differences in brain functional networks. *NeuroImage* 101:681-694.

Pan, W., Kim, J., Zhang, Y., Shen, X. and Wei, P. (2014) A powerful and adaptive association test for rare variants. *Genetics* 197 (4):1081-1095.

Chapter 2

A cautionary note on using secondary phenotypes in neuroimaging genetic studies

2.1 Introduction

Genome-wide association studies (GWASs) have become popular for identifying genetic variants associated with complex diseases and other secondary phenotypes. Most existing GWASs adopt the case-control design, in which a certain number of disease-affected and disease-free individuals are sampled from the corresponding subpopulations respectively [33, 34, 35]. Due to its separate samplings on the subjects conditional on their disease status, a key feature of a case-control sample is that it is not a random sample from the population; though both the case sample and the control sample are a random sample from the corresponding subpopulation, the combined case-control sample is biased for the population because, for example, a fraction of the cases (often close to 50%) much larger than that of the population are included in the case-control sample. Interestingly, when a standard logistic regression model is applied to a case-control sample to assess the disease and a (genetic or other) risk factor association, the case-control sample can be treated as a random sample from the population, though the estimated disease prevalence (i.e. the intercept) is biased Prentice and Pyke [36]. However, when a linear

regression model is applied to other secondary phenotypes to assess their associations with a risk factor, if no adjustment is made for the biased case-control sample, estimation and inference result except under some special situations [14].These conclusions apply to neuroimaging genetic studies. For example, the Alzheimer's Disease Neuroimaging Initiative (ADNI) collected its samples based on participants' disease status: specifically, in ADNI (or more precisely, ADNI-1 as used throughout), 200 healthy controls (HCs),biased 400 subjects with mild cognitive impairment (MCI) and 200 patients with Alzheimer's Disease (AD) were recruited; the set of the ADNI participants is not expected to be a random sample of the age-matched general population. For instance, in the general population, ten to twenty percent of people age 65 or older is known to have mild cognitive impairment (MCI) [5, 6, 7], but nearly a half of the ADNI samples consists of MCI individuals. Hence, although a standard logistic regression model can be applied to draw unbiased inference for genetic associations with the risk of AD, a standard regression model (without any suitable adjustment) may lead to biased inference of genetic associations with secondary phenotypes, such as many neuroimaging phenotypes. On the other hand, surprisingly, to our knowledge, all the publications on analyses of secondary phenotypes for the ADNI data have relied on standard linear regression without any adjustment to or even any discussion on possible problems with the biased ADNI sample [8, 9, 10, 11]. Biased inference may lead to not only biased parameter estimates, but also inflated Type I error rates and reduced power. It is the primary goal of this chapter to address whether such standard linear regression really leads to biased inference for secondary phenotypes using the ADNI data as an example; if so, to what extent.

As a result, findings from previous studies may be questioned. A number of strategies have been proposed for correct inference for secondary phenotypes, including inverse probability weighted regression [12, 13], use of retrospective likelihoods [14, 37, 38] and conditional and other methods [39, 40]. Since the retrospective likelihood method of Lin and Zeng [14] is statistically efficient (but technically more challenging to extend to other more complex situations), while inverse probability weighted regression is easier to implement (but less efficient statistically), we use them as the references against the standard linear regression. In addition, since some imaging genetics studies [41, 42] have considered a variation of the standard linear regression by adjusting for the (primary

phenotype) disease status, we also consider this method.

We first briefly review the above four methods for association analysis of a genetic variant and a secondary phenotype. We then apply the methods to the ADNI data. In Section 2.3.2, realistic simulation studies mimicking the ADNI data are conducted to further investigate possible problems when analyzing a secondary phenotype. Section 2.3.3 provides a simple toy example to demonstrate the problem and offers some intuitive explanations. A summary of our conclusions is given in Section 2.4.

2.2 Methods

Let $\{x_i, Y_i, Z_i, D_i\}$ be the observed data for subject $i = 1, ..., n$, where x_i is an additive genotype score of an SNP of interest, Y_i is a univariate and quantitative secondary phenotype, $D_i = 1$ or 0 is an indicator of the disease (i.e. primary phenotype), and $Z_i = (Z_{i1}, ..., Z_{il})'$ is a vector of covariates. Define the number of controls (with $D_i = 0$) as n_0, and that of cases (with $D_i = 1$) as n_1. A major characteristic of a case-control study is that the disease-status is identified at the beginning of the study, and the sampling of the subjects is conditional on their disease status. One implication is that the combined case-control data may not be a random sample from the population. A proper analysis should take account of the sampling scheme; otherwise biases may result. This chapter considers following four approaches: the first two are standard approaches currently widely used in imaging genetics, while the last two were specifically developed for valid analysis of secondary phenotypes.

2.2.1 Unadjusted linear model

A standard linear model regressing the secondary phenotype (Y_i) on the genotype score (x_i) has been used for testing association between the two:

$$E(Y_i|x_i, Z_i) = \beta_0 + \beta_1 x_i + \beta_z' Z_i, \tag{2.1}$$

and it is assumed that the conditional distribution $f(Y_i|x_i, Z_i)$ is Normal, $N(\beta_0 + \beta_1 x_i + \beta_z' Z_i, \sigma^2)$. Accordingly, based on the likelihood $L = \prod_{i=1}^{n} f(Y_i|x_i, Z_i)$, maximum likelihood is used to draw inference on β_1. For example, as used in the following, the Wald

test is applied to test the null hypothesis $H_0 : \beta_1 = 0$ based on the maximum likelihood estimate (MLE) $\hat{\beta}_1$.

Note that with a case-control sample, in general the above likelihood function $L = \prod_{i=1}^{n} f(Y_i|x_i, Z_i)$ is not appropriate, failing to account for the conditional sampling. Hence, in general the above inference is expected to be biased.

2.2.2 Disease status adjusted linear model

A simple way to adjust for the case-control sampling is to adjust for the disease status (D_i) in a standard linear regression model [41]:

$$E(Y_i|x_i, Z_i, D_i) = \beta_0 + \beta_1 x_i + \beta_z' Z_i + \beta_d D_i,$$

and it is assumed that the distribution density $f(Y_i|x_i, Z_i, D_i)$ is $N(\beta_0 + \beta_1 x_i + \beta_z' Z_i + \beta_d D_i, \sigma^2)$. The likelihood function $L = \prod_{i=1}^{n} f(Y_i|x_i, Z_i, D_i)$ is used for inference of β_1 in the framework of maximum likelihood. Again note that the likelihood $L = \prod_{i=1}^{n} f(Y_i|x_i, Z_i, D_i)$ is in general invalid for the case-control data.

2.2.3 Inverse probability weighted regression

To properly account for biased case-control sampling, a weighted likelihood (or weighted estimating equations) can be used [43, 13, 12]. The weight for each subject is defined to be proportional to the inverse probability of the subject's being sampled into the case-control data. Intuitively, for instance, if the disease is rare in the population, but an equal number of cases and controls are sampled, the weight is used to up-weight the controls and down-weight the affected individuals so that the weighted case-control sample is like a random sample from the population. Monsees et al. [13] discussed such an inverse probability weighted (IPW) regression approach, offering unbiased inference of genotype-secondary phenotype associations, though its statistical efficiency may be low. Following Schifano et al. [12], in this study, the weight (w_i) for subject i was specified as $w_i = p/\pi$ if $D_i = 1$, and $w_i = (1-p)/(1-\pi)$ if $D_i = 0$, where $p = P(D = 1)$ is the disease prevalence in the population, and $\pi = P(D = 1|sampled)$ is the proportion of affected individuals in the case-control sample, which is always substituted with $n_1/(n_0 + n_1)$ throughout. The regression model is the same as equation (2.1), but the likelihood is weighted with w_i, i.e. $L_w = \prod_{i=1}^{n} f(Y_i|x_i, Z_i)^{w_i}$, where $f(Y_i|x_i, Z_i)$

is the density function for a normal distribution, $N(\beta_0 + \beta_1 x_i + \beta_z' Z_i, \sigma^2)$. Maximum likelihood is used for inference. We implemented the above IPW regression approach using geeglm() function in R.

2.2.4 A retrospective likelihood approach

Lin and Zeng [14] and Gosh et al. [38] proposed retrospective likelihoods to properly account for the fact that the case-control data should be conditioned on the disease status. Specifically, a regression model for secondary phenotype data (SPREG) proposed by Lin and Zeng [14] is based on a retrospective likelihood $f(Y_i, x_i, Z_i|D_i) =$

$$\left\{\frac{P(D_i = 1|x_i, Y_i, Z_i)f(Y_i|x_i, Z_i)f(x_i, Z_i)}{P(D_i = 1)}\right\}^{D_i} \times \left\{\frac{P(D_i = 0|x_i, Y_i, , Z_i)f(Y_i|x_i, Z_i)f(x_i, Z_i)}{P(D_i = 0)}\right\}^{1-D_i}, \tag{2.2}$$

where $P(D_i = 1) = \int_Y \int_{x,Z} P(D_i = 1|Y_i, x_i, Z_i)f(Y_i|x_i, Z_i)f(x_i, Z_i)d_Y d_{x,Z}$, $P(D_i = 0) = 1 - P(D_i = 1)$, $P(D_i = 1|Y_i, x_i, Z_i)$ determined by

$$\text{logit}P(D_i = 1|x_i, Y_i, Z_i) = \alpha_0 + \alpha_1 x_i + \alpha_2 Y_i + \alpha_z' Z_i,$$

$f(Y_i|x_i, Z_i)$ is the density function of $N(\beta_0 + \beta_1 x_i + \beta_z Z_i, \sigma^2)$, and $f(x_i, Z_i)$ is treated as nuisance parameters. A profile likelihood is used to eliminate nuisance parameters, which is then maximized by the Newton-Raphson algorithm; maximum likelihood is used to draw inference on β_1. Since this method is likelihood-based, it is efficient. However, due to the presence of high-dimensional nuisance parameters, the (profile) likelihood may be difficult to maximize, leading to some numerical problems as pointed out by Lutz et al. [44] and to be confirmed later, especially if the disease prevalence is unknown or estimated inaccurately [39]. Software for SPREG was downloaded from `http://dlin.web.unc.edu/software/spreg-2/`.

2.3 Results

2.3.1 Real data example

2.3.1.1 Testing with candidate SNPs

We considered a univariate and quantitative secondary phenotype, volume of right hippocampus, for its possible association with each of several SNPs, rs429358, rs2075650,

rs7526034, rs10932886, rs7647307, rs7610017, rs4692256 and rs6463843, which were chosen because they were shown to be highly associated with multiple imaging phenotypes when using the standard (unadjusted) linear regression method [8]. From the ADNI baseline data, we extracted the secondary phenotype, the SNPs and five covariates: gender, education, handedness, age, and intracranial volume (ICV) for association testing.

We regressed hippocampus volume on each genotype score and five covariates using the four methods: standard linear regression without adjusting for disease status (unadj-lm), with adjustment for disease status (D-adj-lm), IPW regression (lm-w), and SPREG [14]. For the latter two methods, an estimate of the AD prevalence in the population is needed, which was obtained based on the following data. In 2014, it was reported that one in nine people age 65 and older (11 percent) had AD, and one third of people age 85 and older (32 percent) had AD; in 2012, 13 percent people age 65 and older were believed to have AD, and nearly half of people age 85 and older had AD [45, 46, 47]. It was not straightforward to determine the disease prevalence, since the AD prevalence for an aging population varies over time and it is not always clear what is the age-matched population based on the given case-control sample. The subjects in our collected data had mean age 75.68 with minimum 56, the first and third quantiles 71.75 and 75.68 respectively. Thus we estimated that the AD prevalence (p) in the population ranged from 0.10 to 0.30. Accordingly we considered disease prevalence $p \in \{0.10, 0.13, 0.16, 0.20, 0.23, 0.27, 0.30\}$, investigating how the results depended on the chosen p. For IPW regression, a subject's weight (w_i) was calculated based on a given p as discussed before; for SPREG, a given p was input to the software program.

We applied the methods to the ADNI data including all $n_0 = 180$ healthy controls (HCs) and $n_1 = 144$ AD patients available from the ADNI baseline data. The results are summarized in Table 2.1. Unadj-lm, lm-w and SPREG suggested significant associations between rs429358/rs2075650 and right hippocampus volume. This is consistent with the results from previous studies [8, 48, 49, 50]. Unadj-lm and SPREG showed more significant p-values. When the disease prevalence p=0.10 or 0.13 was assumed, none of the p-values given by lm-w could reach the genome-wide significance level (5×10^{-8}); however, if p=0.27 or 0.30 was used, rs429358 became highly significant, demonstrating that the results of lm-w were sensitive to the estimate of the disease prevalence p. The

dependence of SPREG on p was to a lesser degree. It is noted that disease adjusted linear model (D-adj-lm) gave no significant p-value for any SNP. Interestingly, when the disease prevalence p=0.23 was assumed, which was reasonable, unadj-lm and SPREG showed p-values close to each other.

To confirm the results in Table 2.1 with a larger sample size, we included additional 311 MCI subjects in the ADNI data. We treated the MCI subjects as controls, and applied the methods with 491 controls and 144 AD patients. In Table 2.2, all p-values became smaller but only rs429358 and rs2075650 showed strong associations with the right hippocampus volume, in agreement to that in Table 2.1. Again, when assuming disease prevalence $p = 0.16$ or 0.20, which was reasonable (because MCIs were treated as controls), unadj-lm, lm-w and SPREG, but not D-adj-lm, all gave similar results.

2.3.1.2 An association scan on chromosome 19

Rather than drawing our conclusions based on only a few SNPs, we conducted a genome-wide scan on chromosome 19. The secondary phenotype was still the right hippocampus volume. We included all the SNPs with minor allele frequency (maf) ≥ 0.05, genotyping rate more than 90%, and surviving the Hardy-Weinberg test with p-value > 0.001, resulting in 9184 SNPs to be tested. Subjects with more than 10% missing genotypes were excluded; only non-Hispanic Caucasians whose right hippocampus volume was measured at baseline were included. As in the previous section, two sample sizes were considered: (1) $n_0 = 180$ controls (HCs) and $n_1 = 144$ AD patients; (2) $n_0 = 491$ controls (including both HCs and MCIs) and $n_1 = 144$ AD patients.

The quantile-quantile (Q-Q) plots in Figures 2.1 and 2.2 show the distributions of the observed p-values against those of the expected (null) p-values. For each method, the pattern shown on the two plots is similar. Surprisingly, both unadj-lm and D-adj-lm had their estimated inflation factors (λ) (almost) 1 in each case, and the observed p-values were in close agreement with the expected ones, suggesting no obvious inflation of their Type I errors. Although the estimated inflation factors for lm-w were also close to 1, there were a few more points falling outside of the confidence regions. Depending on the population disease prevalence p used, the estimated inflation factors of SPREG ranged from 1.06 to 1.16, which were not too bad; however, most strikingly, in every Q-Q plot, there were many observed p-values far more significant than expected, implying

a large portion of likely false positives, presumably due to some numerical problems for those SNPs in SPREG. In our experience, especially for secondary phenotypes with large variances such as brain volumetric measures, SPREG might not converge, and scaling a phenotype by its standard deviation improved its convergence; even with scaling, in this example, SPREG failed to converge for about 1000 SNPs (10%) when the disease prevalence (p) was set to be less than 0.23.

In summary, in an association scan on chromosome 19 with two sample sizes, all methods seemed to give reasonable estimates of inflation factors. In addition, the two unadjusted methods and IPW regression did not show any obvious problem in Type I error inflations; in contrast, SPREG had some numerical problems, giving many SNPs more significant p-values than expected.

2.3.2 Simulations

2.3.2.1 Simulation set-ups

We conducted simulation studies with realistic set-ups to mimic the ADNI data. First we selected two SNPs, rs429358 and rs6463843 in Table 2.1, to represent two association patterns. SNP rs429358 (in gene APOE) is well known for its strong associations with both hippocampus volume and AD [48, 49, 50]; by choosing rs429358, we had a representative case where both the SNP and secondary phenotype are highly associated with the disease risk. On the other hand, rs6463843 (in gene NXPH1) was chosen to reflect an opposite scenario where both the SNP and the secondary phenotype are only moderately associated with the disease. Next, we used the ADNI data to estimated various association parameters for each SNP. Specifically, we fitted a linear regression model with the right hippocampus volume as the secondary phenotype and an SNP (x) and covariates (Z, including gender, education, handedness, age, ICV) as predictors, obtaining the estimated regression coefficients, β_{xy} and β_{zy}. Then a logistic regression model was fitted to determine the effects of SNP (x) and the phenotype (Y) on the disease (D), obtaining the estimated regression coefficients β_{Dy} and β_{Dx}. The parameter values for the two SNPs/set-ups are given in Table 2.3, which were used as the true parameter values for generating simulated data.

To maintain the true correlation structures among the five covariates, we sampled

$Z_i = (Z_{i1}, ..., Z_{i5})$ from the ADNI data in each simulation. An additive genotype score (x_i) was randomly generated from a binomial distribution Bin(2, maf) with maf=0.27 and 0.45 for the two SNPs respectively. The secondary phenotype was generated from a Normal distribution based on the simulated covariates and genotype score $\{Z_i, x_i\}$:

$$Y_i \sim N(\phi \cdot \beta_{xy} x_i + \beta'_{zy} Z_i, \ \ \widehat{\sigma}^2_y), \tag{2.3}$$

where β_{xy} and β_{zy} are presented in Table 2.3, $\widehat{\sigma}^2_y$ was obtained from the sample variance of hippocampus volume, and ϕ is a scaling parameter controlling the association strength between x and Y. When $\phi = 0$, we created a null case with no association; when $\phi = 1$, the effect size was equal to the estimate from the ADNI data.

For each subject i, the disease status D_i was generated from a Bernoulli distribution with probability $P(D_i = 1|x_i, Y_i)$ determined by

$$\text{logit}P(D_i = 1|x_i, Y_i) = \beta_{D0} + \beta_{Dy} Y_i + \beta_{Dx} x_i,$$

where the values of β_{Dy} and β_{Dx} are shown in Table 2.3, and $\beta_{D0} = \text{logit}^{-1} p$. The disease prevalence was set at $p = 0.23$ or 0.10 to mimic that for the ADNI data. Note however that $p = 0.23$ was more reasonable for AD.

To generate a simulated data set, we repeated simulating observations $\{Z_i, x_i, Y_i, D_i\}$ until reaching the predefined sample size of n_1 cases and n_0 controls; any simulated observations not used in the case-control sample were added back to the case-control sample to form a cohort sample. Since a cohort sample was a random sample from the population, while a case-control sample was not, we used the results from cohort samples as benchmarks.

We also used each cohort sample to obtain an estimate $\hat{p}$ of the disease prevalence for the corresponding case-control sample. To investigate the effects of the specified disease prevalence on analysis, three different disease prevalence rates, $\hat{p} - 0.05$, $\hat{p}$, and $\hat{p} + 0.05$, were input to lm-w and SPREG.

For each simulation set-up, the results were based on 10^4 independent simulation replicates.

2.3.2.2 Type I error and power

In Table 2.4, we report the empirical Type I errors for the methods for the null case (with $\phi = 0$). SPREG and lm-w had valid Type I errors in all cases, while the results of

unadj-lm and D-adj-lm largely depended on the simulation set-ups and the true disease prevalence. In set-up 1, where both the SNP and the secondary phenotype were highly associated with the disease risk, unadj-lm, lm-w and SPREG showed proper type I error rates, with the true prevalence $p = 0.23$; however, D-adj-lm gave highly inflated ones. Yet when the true disease prevalence was set at $p = 0.10$, only SPREG had type I errors close to the nominal level (0.05), and lm-w (with $\hat{p}$ applied) gave slightly inflated ones. However, the numerical results suggested that both SPREG and lm-w were sensitive to the pre-specified disease prevalence.

In set-up 2 where the SNP or the secondary phenotype was not highly associated with the disease risk, the Type I error rates of all the methods except D-adj-lm were controlled, though the inflations by D-adj-lm were small to moderate.

In order to ensure the above results were not due to a small sample size, we increased the sample size to $n_0 = n_1 = 500$ and $n_0 = n_1 = 1000$. The empirical Type I error rates of SPREG and lm-w were reliable as compared to unadj-lm and D-adj-lm (not shown). A more extreme disease prevalence $p = 0.01$ (not shown) was also considered, in which the Type I errors of unadj-lm were more inflated, while D-adj-lm performed well as pointed out in Monsees et al. [13].

The empirical power of each method is presented in Table 2.5. In set-up 1, unadj-lm had the highest power (but recall that it had slightly inflated Type I errors), followed by SPREG, then by lm-w. Note the dramatic power loss of D-adj-lm in spite of its severely inflated Type I errors. In set-up 2, D-adj-lm was most powerful but, due to its inflated Type I errors, it should not count; the other three methods were similarly powered.

Figures 2.3 and 2.4 illustrate the distributions of the parameter estimates $\widehat{\beta}_1$ by each method. In set-up 1 (Figure 2.3) lm-w and SPREG provided almost unbiased estimates, while D-adj-lm always yielded largely biased estimates; unadj-lm gave almost unbiased estimates for $p = 0.23$, but slightly biased ones for $p = 0.10$. For set-up 2 (Figure 2.4), only D-adj-lm gave obviously biased estimates.

In all cases, lm-w and SPREG yielded unbiased estimates, while the performance of unadj-lm and D-adj-lm largely depended on the disease prevalence (and simulation set-ups).

In summary, under practical situations mimicking the ADNI data, the standard linear regression method unadj-lm, but not D-adj-lm, performed satisfactorily, giving

results similar to the other two valid methods.

2.3.3 An illustrative example

Finally we used a simple toy example to illustrate the problems with unadj-lm and D-adj-lm. For better visualization, we took a continuous x and no covariate Z; it is easy to see that the main points carry over to the case with a genotype score x and with Z. We assumed a finite population (or equivalently, a random sample from a super-population) containing 9000 controls (with $D = 0$) and 1000 cases (with $D = 1$). For controls, we had a predictor $x \sim N(0, 1)$, while $x \sim N(2, 1)$ for cases. A secondary phenotype Y was distributed as $Y \sim N(2D, 1)$.

Based on the assumed model, we can see that conditional on D, Y was not associated with x, which is confirmed in the left panel of Figure 2.5: for either the control or case group, regressing Y on x yielded a horizontal line; the OLS estimates for the slope parameter of the two groups were 0.005 (SE=0.01) and 0.004 (SE=0.03), respectively. On the other hand, marginally Y was associated with x: the OLS estimate of the slope parameter was 0.260 (SE=0.009).

Now we consider a case-control sample. To minimize the influence of the sampling errors, for simplicity, we took a random sample of 1000 controls and all 1000 cases. As shown in the right panel of Figure 2.5, applying unadj-lm and D-adj-lm led to the OLS estimates of the slope parameter for x as 0.511 (SE=0.019) and 0.006 (SE=0.022) respectively; that is, unadj-lm over-estimated the population marginal association (i.e. 0.511 versus 0.260), while D-adj-lm was on the target for the conditional association (0.006 versus 0) but again off from the marginal association (0.006 versus 0.260). We also applied weighted regression with lm-w: based on the sampling proportions, a weight 9 was assigned to each control and weight 1 to each case in the case-control sample; then we regressed Y on x; the WLS estimate of the slope parameter was 0.279 (SE=0.021), very close to the population marginal association (i.e. 0.279 versus 0.260).

It is simple why unadj-lm may not work for a case-control sample: a case-control sample may not represent the population. More importantly, this example also clearly demonstrates that, even with the data from the whole population (or a large random sample), marginal association based on unadj-lm and conditional association based on D-adj-lm may be quite different. More formally, we are interested in inference for β_1 in

a marginal model

$$E(Y|x) = \beta_0 + \beta_1 x.$$

However, D-adj-lm is based on a conditional model

$$E(Y|x, D) = b_0 + b_1 x + b_2 D,$$

from which we can derive

$$E(Y|x) = E[E(Y|x, D)] = b_0 + b_1 x + b_2 E(D|x).$$

If D is associated with x, say $E(D|x) = a_0 + a_1 x$, then we have

$$E(Y|x) = E[E(Y|x, D)] = b_0 + b_2 a_0 + (b_1 + b_2 a_1)x,$$

based on which we may have $\beta_1 \neq b_1 + b_2 a_1$ unless $b_2 = 0$ or $a_1 = 0$.

2.4 Conclusions

We set out to address whether standard linear regression of secondary phenotypes in a practical neuroimaging genetic study would lead to biased inference, i.e. biased estimates, inflated Type I errors and reduced power. This is an important question given that in general it will lead to biased inference while the current practice in neuroimaging genetics has largely ignored this potential problem. Using the ADNI data as an example, we conducted both real data analyses and simulation studies. Our main conclusion was the following: under practical situations similar to the ADNI data, using standard linear regression without any adjustment (unadj-lm), but not the one adjusting for the disease status (D-adj-lm), to assess SNP-secondary phenotype associations did not appear to cause any severe problem, though cautions still must be taken.

Of course, our main conclusion is only specific to the ADNI data, and is not applicable in general; some general principles were discussed, which might offer some guidelines to practitioners for other applications. The main theoretical reason for our conclusion to hold for the ADNI data (and possibly other data) is due to the high prevalence of the AD (or other disease) in the target population, leading to its small difference from the sampling proportion of the cases in the case-control sample, which is usually close to 50%. In other words, the key issue is how much biased is the case-control sample for the

target population. For example, if the disease is less common, say at 10% in the general population, while as usual about a half of the case-control sample are cases, then a suitable adjustment in analysis is more likely to be necessary. There is also another factor influencing the validity of the standard unadjusted methods: the association strength between the disease and a secondary phenotype. For example, if the disease and the secondary phenotype are not associated, then a standard unadjusted analysis for the secondary phenotype is fine [14]. However, in neuroimaging genetic studies, often a secondary phenotype is of interest simply because it is treated an intermediate phenotype for the disease, suggesting its likely association with the disease. Nevertheless, if the secondary phenotype-disease association is weak, a standard unadjusted analysis of the secondary phenotype may be only slightly biased. We have also discussed why simply adjusting for disease status (D-adj-lm) might not work: in addition to the possible poor representation of a case-control sample for the population, D-adj-lm targets the conditional association between the secondary phenotype and an SNP (after adjusting for possible covariates), not the marginal association of interest, which may be quite different from the conditional association as shown in our toy example in Section 2.3.3.

It is fair to ask why we do not always use one of the valid methods that properly correct for the sampling bias of case-control studies. In this chapter we have considered two representative methods, IPW regression and SPREG; the former is general, more robust [51] and easier to implement but less efficient, while the latter is the opposite. For SPREG, it is challenging to extend it (or other retrospective likelihood methods) to more complex study designs beyond the simple case-control design, such as with longitudinal phenotypes or familial relatedness. Although IPW regression is general and easy to implement, its loss of power may hinder its wide use, especially for small neuroimaging GWASs. In addition, there may be numerical problems with the use of SPREG (see Figures 2.1 and 2.2). Furthermore, both of the methods require an estimate of the disease prevalence in the target population, and their results may be sensitive to the estimate; however, it may not be easy to obtain an accurate estimate, as in the ADNI data, since the target population is not well defined, e.g. with respect to the study participants' age while the AD (or MCI) prevalence largely depends on the age.

Although we have only considered single quantitative secondary phenotype–single SNP associations, we anticipate that our conclusions will be likely to hold for other cases,

such as for binary secondary phenotypes [39, 52], multiple secondary phenotypes [22, 28, 53], longitudinal secondary phenotypes [54, 55], or for gene-gene or gene-environment interactions [11, 56], though further studies are needed.

Table 2.1: P-values for association testing between right hippocampus volume and an SNP with 324 subjects in ADNI.

					lm-w						
SNP	chr	maf	unadj-lm	D-adj-lm	p=0.10	p=0.13	p=0.16	p=0.20	p=0.23	p=0.27	p=0.30
rs429358	19	0.27	7.78e-12	2.03e-02	8.06e-04	7.99e-05	3.63e-06	3.61e-07	3.61e-07	2.02e-09	2.60e-10
rs2075650	19	0.24	5.18e-08	6.11e-02	9.63e-03	2.25e-03	3.42e-04	8.81e-05	8.81e-05	4.77e-06	1.58e-06
rs7526034	1	0.12	6.46e-02	5.68e-01	3.56e-01	2.83e-01	2.17e-01	1.83e-01	1.83e-01	1.32e-01	1.18e-01
rs10932886	2	0.33	3.75e-02	3.31e-01	1.28e-01	9.39e-02	6.62e-02	5.38e-02	5.38e-02	3.93e-02	3.66e-02
rs7647307	3	0.44	1.05e-03	1.33e-01	2.63e-03	1.48e-03	8.19e-04	5.94e-04	5.94e-04	4.02e-04	3.87e-04
rs7610017	3	0.03	4.80e-01	2.11e-02	4.28e-01	4.58e-01	4.93e-01	5.14e-01	5.14e-01	5.43e-01	5.47e-01
rs4692256	4	0.46	2.98e-03	2.87e-02	9.09e-02	5.69e-02	3.21e-02	2.19e-02	2.19e-02	1.02e-02	7.66e-03
rs6463843	7	0.45	6.27e-03	4.65e-02	1.49e-01	9.16e-02	5.11e-02	3.49e-02	3.49e-02	1.72e-02	1.36e-02
			SPREG								
SNP	chr	maf	p=0.10	p=0.13	p=0.16	p=0.20	p=0.23	p=0.27	p=0.30		
rs429358	19	0.27	1.22e-09	2.21e-10	5.42e-11	1.19e-11	4.75e-12	1.83e-12	1.08e-12		
rs2075650	19	0.24	1.44e-06	5.23e-07	2.30e-07	9.59e-08	5.67e-08	3.27e-08	2.40e-08		
rs7526034	1	0.12	1.28e-01	1.12e-01	9.92e-02	8.66e-02	7.95e-02	7.24e-02	6.84e-02		
rs10932886	2	0.33	6.24e-02	5.29e-02	4.67e-02	4.16e-02	3.92e-02	3.71e-02	3.61e-02		
rs7647307	3	0.44	6.11e-03	4.02e-03	2.85e-03	1.97e-03	1.59e-03	1.27e-03	1.12e-03		
rs7610017	3	0.03	2.15e-01	2.58e-01	2.97e-01	3.42e-01	3.70e-01	4.02e-01	4.22e-01		
rs4692256	4	0.46	2.65e-03	2.55e-03	2.50e-03	2.47e-03	2.45e-03	2.43e-03	2.42e-03		
rs6463843	7	0.45	6.28e-03	5.78e-03	5.51e-03	5.35e-03	5.32e-03	5.33e-03	5.36e-03		

notes: maf based on the 324 subjects

Table 2.2: P-values for association testing between right hippocampus volume and an SNP when subjects with MCI were included in ADNI.

					lm-w						
SNP	chr	maf	unadj-lm	D-adj-lm	p=0.10	p=0.13	p=0.16	p=0.20	p=0.23	p=0.27	p=0.30
rs429358	19	0.31	8.12e-16	2.52e-11	2.22e-15	2.22e-16	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
rs2075650	19	0.26	1.21e-07	4.61e-05	1.31e-06	3.99e-07	1.46e-07	5.23e-08	5.23e-08	2.12e-08	2.00e-08
rs7526034	1	0.13	1.19e-03	4.71e-03	2.68e-03	2.29e-03	2.12e-03	2.16e-03	2.16e-03	2.97e-03	3.69e-03
rs10932886	2	0.33	1.26e-01	3.42e-01	2.31e-01	1.88e-01	1.58e-01	1.31e-01	1.31e-01	1.11e-01	1.09e-01
rs7647307	3	0.44	1.43e-03	1.89e-02	1.26e-03	1.05e-03	9.60e-04	9.87e-04	9.87e-04	1.44e-03	1.85e-03
rs7610017	3	0.04	9.36e-01	7.23e-01	7.49e-01	8.15e-01	8.87e-01	9.90e-01	9.90e-01	8.24e-01	7.46e-01
rs4692256	4	0.46	8.15e-04	4.52e-03	7.06e-03	3.94e-03	2.28e-03	1.18e-03	1.18e-03	4.74e-04	3.55e-04
rs6463843	7	0.47	7.87e-04	1.10e-03	1.14e-03	8.94e-04	7.60e-04	6.87e-04	6.87e-04	7.33e-04	8.03e-04
			SPREG								
SNP	chr	maf	p=0.10	p=0.13	p=0.16	p=0.20	p=0.23	p=0.27	p=0.30		
rs429358	19	0.31	5.33e-15	1.55e-15	4.44e-16	2.22e-16	0.00e+00	0.00e+00	0.00e+00		
rs2075650	19	0.26	6.98e-07	3.41e-07	1.90e-07	1.02e-07	7.03e-08	4.72e-08	3.72e-08		
rs7526034	1	0.13	1.50e-03	1.33e-03	1.21e-03	1.11e-03	1.06e-03	1.00e-03	9.72e-04		
rs10932886	2	0.33	1.82e-01	1.61e-01	1.46e-01	1.31e-01	1.22e-01	1.14e-01	1.09e-01		
rs7647307	3	0.44	3.87e-03	2.74e-03	2.05e-03	1.49e-03	1.22e-03	9.93e-04	8.77e-04		
rs7610017	3	0.04	8.54e-01	8.79e-01	9.00e-01	9.23e-01	9.37e-01	9.53e-01	9.63e-01		
rs4692256	4	0.46	1.15e-03	9.67e-04	8.53e-04	7.59e-04	7.14e-04	6.73e-04	6.51e-04		
rs6463843	7	0.47	6.67e-04	6.55e-04	6.57e-04	6.74e-04	6.94e-04	7.27e-04	7.55e-04		

notes: maf based on the 635 subjects

Table 2.3: Simulation set-ups: parameter values used.

Set-up	SNP	maf	p	ϕ	β_{xy}	β_{Dx}	p-value	β_{Dy}	p-value
1	rs429358	0.27	0.23, 0.10	$0 \leq \phi \leq 1$	-270	1.76	4.61e-16	-4.6e-04	2.29e-15
2	rs6463843	0.45	0.23, 0.10	$0 \leq \phi \leq 2$	-106	0.61	8.66e-05	-3.3e-04	4.58e-08

Figure 2.1: Q-Q plots of the p-values for each method when applied to chromosome 19 in ADNI.

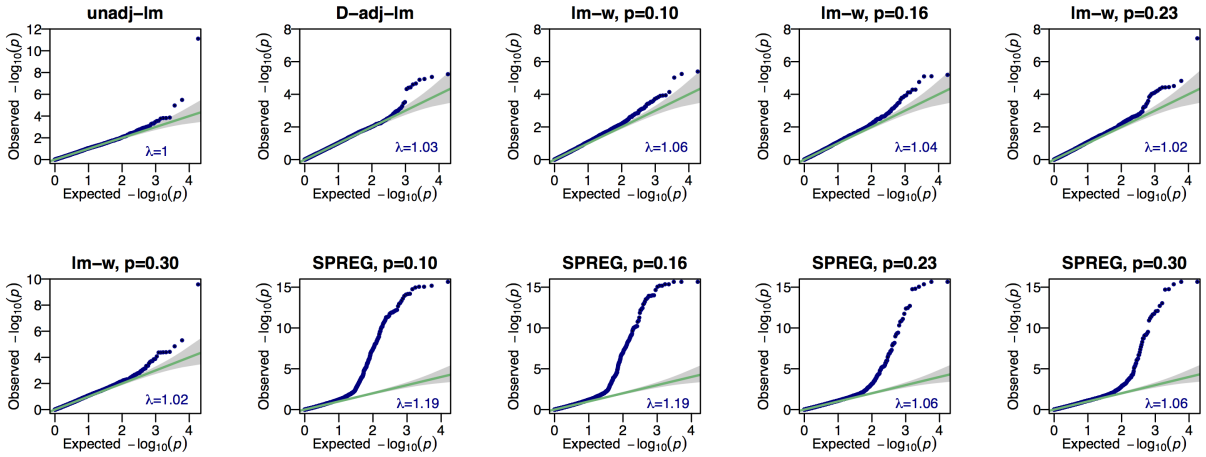


Figure 2.2: Q-Q plots of the p-values for each method when applied to chromosome 19 and subjects with MCI were included in ADNI.

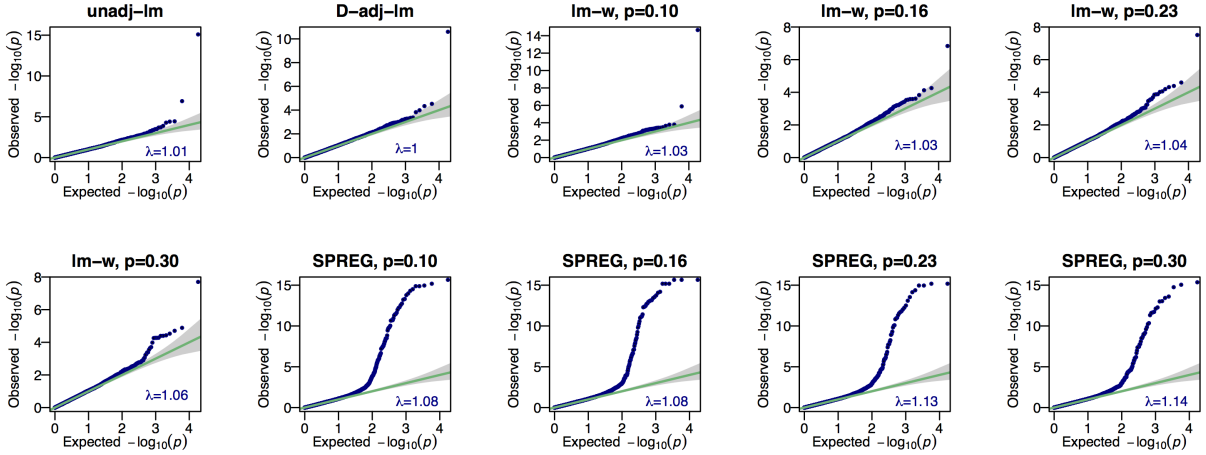


Table 2.4: Type I error rates based on 10^4 simulations with sample size 324.

						lm-w			SPREG		
Set-up	p	α	cohort	unadj-lm	D-adj-lm	$\widehat{p}-0.05$	$\widehat{p}$	$\widehat{p}+0.05$	$\widehat{p}-0.05$	$\widehat{p}$	$\widehat{p}+0.05$
1	0.23	0.050	0.0494	0.0565	0.1457	0.0573	0.0546	0.0550	0.0529	0.0541	0.0545
		0.010	0.0102	0.0101	0.0484	0.0120	0.0124	0.0128	0.0125	0.0117	0.0111
		0.005	0.0057	0.0051	0.0287	0.0056	0.0059	0.0059	0.0060	0.0058	0.0051
		0.001	0.0016	0.0009	0.0099	0.0014	0.0010	0.0009	0.0011	0.0008	0.0009
	0.10	0.050	0.0498	0.0747	0.1103	0.0671	0.0614	0.0621	0.0626	0.0561	0.0601
		0.010	0.0114	0.0188	0.0328	0.0166	0.0133	0.0131	0.0149	0.0124	0.0139
		0.005	0.0059	0.0109	0.0198	0.0078	0.0065	0.0076	0.0083	0.0062	0.0072
		0.001	0.0010	0.0030	0.0055	0.0017	0.0012	0.0015	0.0017	0.0014	0.0016
2	0.23	0.050	0.0492	0.0482	0.0837	0.0534	0.0543	0.0544	0.0540	0.0538	0.0532
		0.010	0.0091	0.0104	0.0221	0.0116	0.0107	0.0113	0.0109	0.0109	0.0111
		0.005	0.0051	0.0044	0.0114	0.0064	0.0057	0.0057	0.0061	0.0058	0.0056
		0.001	0.0009	0.0006	0.0028	0.0011	0.0014	0.0015	0.0010	0.0009	0.0009
	0.10	0.050	0.0492	0.0537	0.0698	0.0549	0.0549	0.0560	0.0562	0.0548	0.0557
		0.010	0.0085	0.0115	0.0177	0.0129	0.0128	0.0126	0.0132	0.0130	0.0130
		0.005	0.0048	0.0063	0.0113	0.0067	0.0070	0.0072	0.0070	0.0071	0.0071
		0.001	0.0013	0.0012	0.0034	0.0014	0.0018	0.0016	0.0023	0.0020	0.0019

Table 2.5: Power based on 10^4 simulations with sample size 324.

						lm-w			SPREG		
Set-up	p	ϕ	cohort	unadj-lm	D-adj-lm	$\widehat{p}-0.05$	$\widehat{p}$	$\widehat{p}+0.05$	$\widehat{p}-0.05$	$\widehat{p}$	$\widehat{p}+0.05$
1	0.23	0.10	0.1240	0.1029	0.0722	0.0712	0.0829	0.0937	0.0775	0.0861	0.0949
		0.30	0.6567	0.3969	0.0676	0.2571	0.3078	0.3520	0.3172	0.3528	0.3749
		0.50	0.9744	0.7829	0.2574	0.6069	0.6811	0.7328	0.7138	0.7444	0.7654
		0.70	0.9996	0.9622	0.5924	0.8747	0.9184	0.9437	0.9414	0.9514	0.9567
		1.00	1.0000	0.9993	0.9377	0.9931	0.9976	0.9985	0.9987	0.9990	0.9992
	0.10	0.10	0.1723	0.1690	0.0619	0.0694	0.0878	0.1062	0.0708	0.0902	0.1129
		0.30	0.8540	0.5470	0.1088	0.1969	0.2744	0.3533	0.2937	0.3803	0.4381
		0.50	0.9982	0.8794	0.3607	0.4688	0.6018	0.7119	0.6895	0.7729	0.8166
		0.70	1.0000	0.9876	0.7083	0.7628	0.8729	0.9270	0.9340	0.9638	0.9757
		1.00	1.0000	1.0000	0.9702	0.9659	0.9920	0.9980	0.9963	0.9996	0.9999
2	0.23	0.30	0.1577	0.0961	0.1874	0.1066	0.1025	0.0999	0.1101	0.1057	0.1026
		0.60	0.4817	0.2462	0.3570	0.2642	0.2639	0.2611	0.2757	0.2664	0.2606
		1.00	0.8815	0.5688	0.6262	0.5698	0.5784	0.5816	0.6000	0.5900	0.5844
		1.50	0.9960	0.8877	0.8827	0.8860	0.8917	0.8950	0.9020	0.8973	0.8942
		2.00	1.0000	0.9887	0.9770	0.9865	0.9882	0.9888	0.9897	0.9896	0.9895
	0.10	0.30	0.2020	0.0847	0.1707	0.1028	0.0975	0.0949	0.1144	0.1062	0.1001
		0.60	0.5938	0.2350	0.3429	0.2387	0.2430	0.2448	0.2899	0.2763	0.2662
		1.00	0.9433	0.5642	0.6155	0.5241	0.5417	0.5571	0.6216	0.6090	0.5990
		1.50	0.9992	0.8925	0.8833	0.8533	0.8719	0.8865	0.9169	0.9117	0.9073
		2.00	1.0000	0.9909	0.9787	0.9800	0.9857	0.9890	0.9929	0.9927	0.9925

Figure 2.3: Simulation set-up 1: Distributions of the estimates $\widehat{\beta}_1$ from each method with two different values of the disease prevalence p.

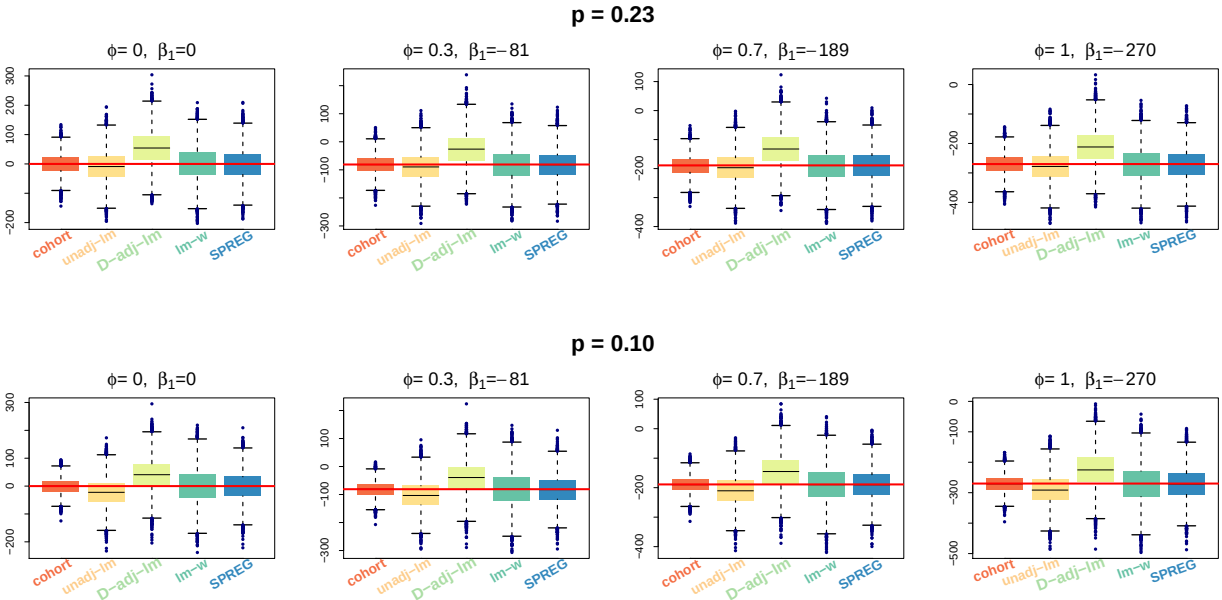


Figure 2.4: Simulation set-up 2: Distributions of the estimates $\widehat{\beta}_1$ from each method with two different values of the disease prevalence p.

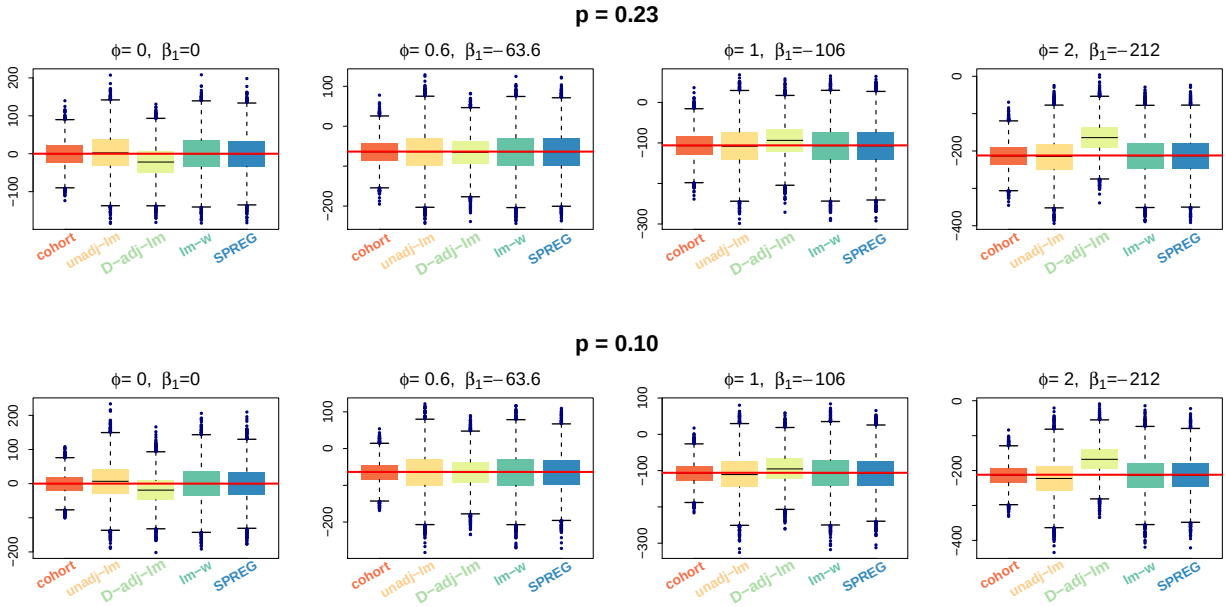
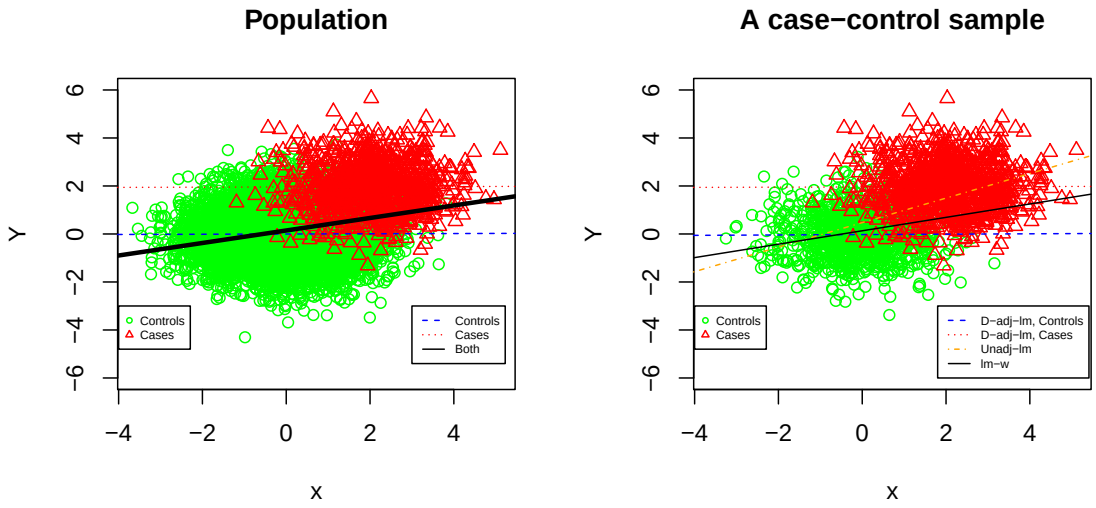


Figure 2.5: An illustrative example. The left panel is for a population with 9000 controls and 1000 cases, while the right panel is for a case-control sample with 1000 controls and 1000 cases.



Chapter 3

Adaptive testing for multiple traits in a proportional odds model with applications to detect SNP-brain network associations

3.1 Introduction

Imaging genetics leverages the strengths of both neuroimaging and genetic studies. In imaging genetic studies, in addition to genotypic data, hundreds to thousands of neuroimaging and neuropsychological phenotypes are collected as intermediate phenotypes. The use of intermediate phenotypes provides some advantages over that of a disease status, both in improving power for discovering risk genes and in understanding underlying pathogenic mechanisms of neurological disorder like Alzheimer's disease (AD) [1, 2]. Given typically small effect sizes of common genetic variants and mounting expenses in increasing sample sizes, it is always of interest to develop more powerful and flexible statistical tests; in particular, in neuroimaging genetic studies, one may want to take advantage of and incorporate synchronous brain activities in multiple brain regions by using multiple imaging traits.

Although many existing methods have appeared in practical application, [12, 15,

20, 28, 57, 58], association analyses for multiple phenotypes are challenging, because a uniformly most powerful test does not exist. A key issue in multi-trait analysis is how to maximize the statistical power in the presence of many non-associated traits, while gaining the power when many or most of the traits are weakly associated with the SNP of interest. In the former situation, one can avoid losing testing power by utilizing only few top associated traits as in the minP test or TATES [15], or using principal components analysis (PCA), principal components of heritability (PCH) or related methods for dimension reduction [18, 19, 20, 21, 22, 23, 24]. In contrast, in the latter situation with many weak associations, jointly analyzing multiple traits by aggregating their weak effects together is necessary, as done in the burden tests [25, 8] and variance component tests [26]. Yet the true association pattern is unknown in practice, and a statistical method has to be flexible enough to adapt to the given data; it would be desirable for a test to capture joint associations of multiple traits with dense association signals while to maintain high statistical power even with sparse association patterns. For example, Zhang et al. [28] proposed a family of association tests, so-called sum of powered score (SPU) tests, and its adaptive version, adaptive SPU (aSPU) test. An SPU(γ) test employs a positive integer γ to incorporate the use of weights to be powerful for a certain association pattern (e.g. the proportion of associated traits with the SNP of interest). A larger γ upweights the traits more highly associated with the SNP, in which way the test's power remains high even in the presence of many non-associated traits. Since the true association pattern is unknown, the aSPU test is proposed to combine information across multiple SPU(γ) tests, each targeting a possible true association pattern. Accordingly, the aSPU test chooses γ and thus weights based on the data so that it can maintain high statistical power in a wide range of scenarios. As in many existing approaches, Zhang et al. [28] assumed a large sample setting, in which the number of phenotypes (p) is much smaller than the sample size (n), and treated the additive genotype score as a continuous predictor and the multiple phenotypes as correlated responses in a generalized estimating equation (GEE) framework. Among others, as shown by Wang [57], the use of the additive inheritance model may lead to loss of power when the assumption is violated.

In this chapter, we propose a new adaptive test built on a proportional odds model (POM), in which the genotype score (i.e. 0, 1, 2 as the minor allele count) is treated as

an ordinal categorical response while the multiple phenotypes as the predictors. Suppose we group subjects by their genotype score values. The POM [59] assumes that there is a continuous unmeasured latent variable whose values determine the observed ordinal value (i.e. genotype score), and the cut-points of the latent variable are envisaged as an intercept in a cumulative logit of the ordinal category. The model assumes identical log-odds ratios across cumulative logits, but the intercept depends on the category, which allows a non-linear relationship between the genotype and the phenotypes. In addition, the model is flexible in that different types of phenotypes (e.g. quantitative or discrete ones) can be equally employed as predictors. Although POMs have been used in association testing for multiple traits [58, 57, 60], we differ from the above works in developing an adaptive test, which, in contrary to that of existing works, can be applied to a high dimensional setting where the number of the traits (p) can be much larger than the sample size (n), as well as to a usual small p setting. Often high dimensional traits are of interest, for which most existing approaches focus on reducing the dimension of the traits, e.g. by a screening procedure, independent component analysis (ICA), canonical correlation analysis (CCA), PCA, or sparse regression [21, 22, 23, 24, 61, 62]. Yet a dimension reduction approach may lose power, because it is likely to ignore weakly associated traits or still include non-associated traits. Given that common variants have weak effects, and multiple phenotypes are prone to be correlated in measuring the same underlying biological trait, often weak effects accumulate for an overall association. Compared to this limitation, the proposed method has been developed in identifying SNPs with pleiotropic effects on multiple traits in a different context with GEE [28].

A set of brain measures from multiple regions of interest (ROIs), or brain circuits including structural or functional connectivity between multiple ROIs, can be the phenotypes of interest. As MRI driven phenotypes, ROI level cortical gray matter thicknesses, surface areas, volumes, or amyloid measurements [8, 24, 63] are widely used. A number of papers have studied genetic effects on brain connectivity; most focused on the analyses for candidate genes or heritability, and used connectivity phenotypes estimated by ICA. [61, 62, 64, 65, 66]. A graph model provides a framework for functional or structural connectivity; between any two ROIs as two nodes in the graph, a pairwise association based on their temporal correlations of BOLD signals or on the total number of fibers interconnecting them is used for their functional or structural connectivity

[67, 68]; for r ROIs, we have $r \times (r-1)/2$ connections, as connectivity phenotypes. Bringing more complex imaging phenotypes such as brain networks to the large scale genetic studies is also considered [69, 70]. Several studies conducted GWAS for brain connectivity analyses, but used only single connectivity [69, 71], while it may be more fruitful to simultaneously exploit multiple phenotypes for a whole network.

We will demonstrate the promising performance of the new test with both real data and simulated data. The new test was applied to Alzheimer's Disease Neuroimaging Initiative (ADNI) data to identify multi trait-single SNP associations. We focus on brain measures in the ROIs for default mode network (DMN), partly because DMN can be used as a clinical diagnostic indicator for Alzheimer's disease [61, 62, 72]. In particular, cortical gray matter (GM) thicknesses from DMN ROIs were employed for its capability of detecting preclinical Alzheimer's disease [73]. In addition, we considered functional connectivity in DMN as multiple phenotypes, which are useful but under-utilized in previous studies. The application of the new method led to the identification of several top SNPs of biological interest. In the simulation studies, we demonstrate that the proposed method showed performance competitive to GEE-based ones [28] and potential power gains when the genetic inheritance mode was non-additive but dominant.

In the following, we introduce the new adaptive test in a POM in Section 3.2. In Sections 3.3, the new method and Zhang et al. [28] are compared with applications to the ADNI data and simulated data. We end with a short summary of the conclusions and future directions in Section 3.4.

3.2 Methods

3.2.1 A proportional odds model

Suppose subject i has a genotype score $Y_i = 0, 1$ or 2 (i.e. count of the minor allele) for a SNP of interest; Y_i indicates $J = 3$ ordered categories. We observe p multiple phenotypes $X_i = (x_{i1}, ..., x_{ip})$ and l covariates $Z_i = (z_{i1}, ..., z_{il})$ for $i = 1, ..., n$. Denote $n_j = \sum_{i=1}^{n} I(Y_i = j)$ and $\pi_j = Pr(Y_i = j)$ for the j ordered category.

First we describe the POM, which is widely used for ordinal response data [57, 58, 59]. Define two sets of regression coefficient vectors: $\beta = (\beta_1, ..., \beta_p)'$ and $\delta = (\delta_1, ..., \delta_l)'$,

and a vector of intercepts $\alpha = (\alpha_0, ..., \alpha_{J-2})'$. The cumulative logit model becomes

$$\text{logit}[Pr(Y_i \leq j)] = \alpha_j + Z_i\delta + X_i\beta \quad \text{for } j = 0, ..., J-2 \tag{3.1}$$

This model assumes that Z or X have identical effects across $(J-1)$ cumulative logit models (i.e. δ and β) but the intercepts α_j vary with j with constraints $\alpha_0 < \alpha_1 < \cdots < \alpha_{J-2}$. A likelihood for equation (3.1) can be derived based on the multinomial distribution for the categorical variable Y_i. The $(J+l+p-1)$ dimensional score vector for POM in equation (3.1) is a gradient of the log likelihood with respect to $\theta = (\alpha', \delta', \beta')'$: $U_\theta = (U_\alpha', U_\delta', U_\beta')'$.

Consider a single multinomial observation $(n_0, ..., n_{J-1})$. The likelihood function is $\pi_0^{n_0} \cdots \pi_{J-1}^{n_{J-1}}$. McCullagh (1980) re-parametrized the likelihood in terms of a cumulative probability $r_k = \sum_{j=0}^{k} \pi_j$. Simplifying McCullagh's [59] results, we arrive at a closed form for each component of the score, U_α, U_δ and U_β, as follows. Define $R_k = \sum_{j=0}^{k} n_j$ and $S_k = R_k/n$ where $R_{J-1} = n$ and $S_{J-1} = 1$. The log likelihood can be written as the sum of $J-1$ quantities

$$l = n\Big[\{S_0\phi_0 - S_1 g(\phi_0)\} + \{S_1\phi_1 - S_2 g(\phi_1)\} + \cdots + \{S_{J-2}\phi_{J-2} - g(\phi_{J-2})\} \Big],$$

where $\phi_j = \text{logit}(r_j/r_{j+1})$ and $g(\phi_j) = \log\{r_{j+1}/(r_{j+1} - r_j)\}$. The cumulative logit model in equation (3.1) is rewritten as $\text{logit}(r_j) = H_{i(j)}\theta$ with $\theta = (\alpha', \delta', \beta')'$ and $H_{i(j)} = (0, ..., 1, ..., 0, Z_i, X_i)$ where the 1 occurs in the position $j+1$ where $j \in \{0, ..., J-2\}$. Denote θ_w and $H_{i(j,w)}$ are the wth element of the vector θ and $H_{i(j)}$ respectively.

By using the chain rule, the gradient of the log-likelihood with respect to $\theta = (\alpha', \delta', \beta')'$ is obtained as,

$$\begin{aligned}
\frac{\partial l}{\partial \theta_w} &= \sum_{j=0}^{J-2} \frac{\partial l}{\partial \phi_j}\frac{\partial \phi_j}{\partial r_j}\frac{\partial r_j}{\partial \theta_w}, \\
\frac{\partial l}{\partial \phi_j} &= R_j - R_{j+1}\frac{r_j}{r_{j+1}}, \\
\frac{\partial \phi_j}{\partial r_j} &= r_{j+1}/\{r_j(r_{j+1} - r_j)\}, \\
\frac{\partial r_j}{\partial \theta_w} &= r_j(1 - r_j)H_{i(j,w)} - r_j(1 - r_{j+1})H_{i(j+1,w)}.
\end{aligned}$$

Considering individual level probabilities $r_{ij} = Pr(Y_i \leq j|X_i, Z_i)$ and $\phi_{ij} = \text{logit}\{r_{ij}/r_{i(j+1)}\}$, the score vector is defined by

$$U_\theta = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \theta} = \sum_{i=1}^{n} \sum_{j=0}^{J-2} \frac{\partial l_i}{\partial \phi_{ij}} \frac{\phi_{ij}}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \theta}.$$

Each component of $U_\theta = \left(U'_\alpha,\ U'_\delta,\ U'_\beta\right)'$ arrives

$$\begin{aligned}
U_\beta &= \sum_{i=1}^{n} \sum_{j=0}^{J-2} (1 - r_{i(j-1)} - r_{ij}) \cdot I(Y_i = j) \cdot X_i, \\
U_\delta &= \sum_{i=1}^{n} \sum_{j=0}^{J-2} (1 - r_{i(j-1)} - r_{ij}) \cdot I(Y_i = j) \cdot Z_i, \\
U_{\alpha j} &= \sum_{i=1}^{n} \frac{1 - r_{ij}}{r_{i(j+1)} - r_{ij}} \Big[I(Y_i \leq j) r_{i(j+1)} - I\{Y_i \leq (j+1)\} r_{ij} \Big] \\
&\quad - \frac{1 - r_{ig}}{r_{ig} - r_{i(g-1)}} \Big[I\{Y_i \leq (j-1)\} r_{ij} - I(Y_i \leq j) r_{i(j-1)} \Big],
\end{aligned}$$

with $j \in \{0, ..., J-2\}$.

We estimate the covariance matrix of the score vector $Cov(U_\theta)$ based on the observed Fisher information matrix:

$$\begin{aligned}
Cov(U_\theta) &= [A_{ws}], \\
A_{ws} &= \sum_{i=1}^{n} \sum_{j=0}^{J-2} \frac{r_{i(j+1)}}{r_{ij}(r_{i(j+1)} - r_{ij})} q_{jw}(i) q_{js}(i), \\
q_{jw}(i) &= r_{ij}(1 - r_{ij}) H_{i(j,w)} - r_{ij}(1 - r_{i(j+1)}) H_{i(j+1,w)}.
\end{aligned}$$

The covariance matrix can be partitioned according to the parameter components $(\alpha', \delta')'$ and β' into $Cov(U_\theta) = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$.

Specifically, to test the association between multiple phenotypes and the genotype score, one can test the null hypothesis $H_0 : \beta = (\beta_1, ..., \beta_p)' = 0$ using the score vector

$$U_\beta = \sum_{i=1}^{n} \sum_{j=0}^{J-2} (1 - \hat{r}_{i(j-1)} - \hat{r}_{ij}) \cdot I(Y_i = j) \cdot X_i, \tag{3.2}$$

where $\widehat{r}_{ij} = exp(\widehat{\alpha}_j + Z_i\widehat{\delta})/[1 + exp(\widehat{\alpha}_j + Z_i\widehat{\delta})]$ for $j = 0$ or 1 is from the fitted null model of equation (3.1); $\widehat{\alpha}_j$ and $\widehat{\delta}$ can be estimated by a numerical procedure (e.g. Fisher scoring or Newton-Raphson) as implemented in R package `MASS` or `VGAM`.

Under H_0: $\beta = 0$, U_β asymptotically follows a multivariate normal distribution, $\mathcal{MN}(0, \Sigma_\beta)$, with $\Sigma_\beta = V_{22} - V_{21}V_{11}^{-1}V_{12}$, in which the estimates $\widehat{\alpha}_j$ and $\widehat{\delta}$ are used. For ease of notation, we suppress β and take $U = U_\beta$ and $\Sigma = \Sigma_\beta$ hereafter.

As a global test, the score test has been widely considered [12, 57]. The score test statistic for testing H_0 is

$$\text{Score} = U'\Sigma^{-1}U,$$

which follows a chi-squared distribution with p degrees of freedom. The simplicity of the score test is convenient, but comes at a potential cost with p degrees of freedom.

3.2.2 An adaptive test

Suppose U_k is the kth component of the score vector $U = (U_1, ..., U_p)'$. The SPU(γ) test statistic is defined as

$$\text{SPU}(\gamma) = \sum_{k=1}^{p} U_k^\gamma$$

for an integer $\gamma \geq 1$. The SPU(γ) test can be considered as a weighted score test [25] with weights $U_k^{\gamma-1}$ on each component k. SPU(1) and SPU(2) are similar to a burden test and a variance-component score test (i.e. kernel machine regression) respectively [74, 75]. As the parameter γ increases, the SPU(γ) test puts higher weights on the traits with larger $|U_k|$, those more strongly associated traits. Accordingly, if the truly associated traits are sparse, using a larger γ would offer higher power. For an extreme situation, as $\gamma \to \infty$ as an even integer, it only takes the maximum component of the score vector and the test statistic is defined as $\text{SPU}(\infty) = \max_{k=1}^{p} |U_k|$, which is closely related to the UminP test (if varying variances of U_k's are ignored). In practice, because it is unknown which γ value would yield high power, an adaptive SPU (aSPU) test is introduced to combine the evidence across multiple SPU tests:

$$\text{aSPU} = \min_{\gamma \in \Gamma} P_{\text{SPU}(\gamma)}$$

where $P_{\text{SPU}(\gamma)}$ is the p-value of SPU(γ), and Γ is a set for candidate integer $\gamma \geq 1$; $\Gamma = \{1, 2, ..., 8, \infty\}$ was used for its good performance in all numerical studies.

If the sample size is large enough for the asymptotic null distribution of the score vector to hold, we use a simulation method to estimate the p-values of all the SPU and aSPU

tests. A large number of the null score vectors can be generated from the null distribution: $U^{(b)} \sim \mathcal{MN}(0, \Sigma)$ for $b = 1, ..., B$. Then the null statistics $\text{SPU}(\gamma)^{(b)}$ are obtained for each b. The p-value of each $\text{SPU}(\gamma)$ is calculated as $P_{\text{SPU}(\gamma)} = \big[\sum_{b=1}^{B} I(|\text{SPU}(\gamma)^{(b)}| \geq |\text{SPU}(\gamma)|) + 1\big]/(B+1)$, where $I(\cdot)$ denotes the indicator function. Based on the same set of null statistics, at the same time, we calculate the p-value for the aSPU test as $P_{\text{aSPU}} = \big[\sum_{b=1}^{B} I(\text{aSPU}^{(b)} \leq \text{aSPU}) + 1\big]/(B+1)$ where $\text{aSPU}^{(b)} = \min_{\gamma \in \Gamma} P_{\gamma}^{(b)}$ and $P_{\gamma}^{(b)} = \big[\sum_{b_1 \neq b} I|\text{SPU}(\gamma)^{(b_1)}| \geq |\text{SPU}(\gamma)^{(b)}| + 1\big]/B$.

Yet when the sample size is not large as compared to p, the asymptotic null distribution of the score vector may not hold. Accordingly, we use a permutation method to estimate the p-values of all the tests. A benefit of using the permutation method is that we do not need to estimate Σ; for a large p, $\Sigma_{p \times p}$ could be singular or unstable. The null score vector $U^{(b)}$ can be generated by permuting subject indices for the phenotypes: suppose that $\{1, 2, ..., n\}$ is permuted to $\{\sigma(1), \sigma(2), ..., \sigma(n)\}$; replace X_i in equation (3.2) with $X_{\sigma(i)}$. With the null score $U^{(b)}$ obtained from each permutation b, the null statistics $\text{SPU}(\gamma)^{(b)}$ are computed for each γ. The p-values of each $\text{SPU}(\gamma)$ and aSPU are calculated as before.

To distinguish the proposed tests from those based on GEE, we call the proposed tests POM-based; if needed, we will use notation such as POM-aSPU.

3.2.3 A doubly adaptive test

Suppose we consider p connectivity phenotypes as a brain network. Due to a large number of parameters in estimating a network, often a penalized method is apply to strike a good bias-variance trade-off in the resulting estimate [21, 22, 76]. A simple approach is to regularize a network estimate through hard thresholding: given an unregularized estimate X_i and a given threshold t, a regularized estimate is $X_i(t) = X_i \circ I(|X_i| > t)$, where $\circ$ represents an element-wise product. At each threshold t, the model and score vector are re-written as

$$\begin{aligned} \text{logit}[Pr(Y_i \leq j)] &= \alpha_j + Z_i\delta + X_i(t) \cdot \beta, \quad j = 0,\ 1, \\ U(t) &= \textstyle\sum_{i=1}^{n} \sum_{j=0}^{J-2} (1 - \hat{r}_{i(j-1)} - \hat{r}_{ij}) \cdot I(Y_i = j) \cdot X_i(t). \end{aligned}$$

To adapt two parameters t and γ, we employ a doubly adaptive test with the statistics:

$$\begin{aligned}\text{SPU}(t,\gamma) &= \sum_{k=1}^{p}[U_k(t)]^\gamma,\\ \text{aSPU}(\gamma) &= \min_t P_{\text{SPU}(t,\gamma)},\\ \text{daSPU} &= \min_\gamma P_{\text{aSPU}(\gamma)},\end{aligned}$$

where $U_k(t)$ is the kth element of $U(t)$. P-values of $\text{SPU}(t,\gamma)$ and $\text{aSPU}(\gamma)$, daSPU tests can be obtained similarly as before, based on the same set of simulated or permuted null scores $U^{(b)}$ for $b = 1, ..., B$. The procedure is described below:

Step 0. Obtain the null scores $U^{(b)}$ using either simulations or permutations.

Step 1. From the null scores $U^{(b)}$, obtain $U(t)^{(b)}$ with candidate thresholds t's, and the null statistics $\text{SPU}(t,\gamma)^{(b)}$ for each γ's and t's.

Step 2. From the null statistics $\text{SPU}(t,\gamma)^{(b)}$, obtain the null statistics $\text{aSPU}(\gamma)^{(b)}$:
$P^{(b)}_{t,\gamma} = \big[\sum_{b_1\neq b} I|\text{SPU}(t,\gamma)^{(b_1)}| \geq |\text{SPU}(t,\gamma)^{(b)}| + 1\big]/B$,
$\text{aSPU}(\gamma)^{(b)} = \min_t P^{(b)}_{t,\gamma}$.

Step 3. From the null statistics $\text{aSPU}(\gamma)^{(b)}$, obtain the null statistics $\text{daSPU}^{(b)}$:
$P^{(b)}_{\gamma} = \big[\sum_{b_1\neq b} I[|\text{aSPU}(\gamma)^{(b_1)}| \leq |\text{aSPU}(\gamma)^{(b)}| + 1\big]/B$,
$\text{daSPU}^{(b)} = \min_\gamma P^{(b)}_{\gamma}$.

Step 4. Based on the above null statistics, the p-values of $\text{SPU}(t,\gamma)$, $\text{aSPU}(\gamma)$, daSPU tests are obtained:

$$P_{\text{SPU}(t,\gamma)} = \big[\textstyle\sum_{b=1}^{B} I(|\text{SPU}(t,\gamma)^{(b)}| \geq |\text{SPU}(t,\gamma)|) + 1\big]/(B+1),$$

$$P_{\text{aSPU}(\gamma)} = \big[\textstyle\sum_{b=1}^{B} I(|\text{aSPU}(\gamma)^{(b)}| \leq |\text{SPU}(\gamma)|) + 1\big]/(B+1),$$

$$P_{\text{daSPU}} = \big[\textstyle\sum_{b=1}^{B} I(|\text{daSPU}^{(b)}| \leq |\text{daSPU}| + 1\big]/(B+1).$$

3.2.4 Comparison with existing tests

As to be shown, several tests (e.g. score or aSPU) based on POM often give similar results with the corresponding GEE-based tests; this can be explained by the closeness of the score vector of POM and that of GEE with a working independence model.

Denote $n_j = \sum_{i=1}^{n} I(Y_i = j)$ as the genotype group size. Without any covariate and with a 3-categorical Y_i, each score vector can be shown as Zhang et al. [28]

$$U_{GEE} = \frac{-n_1 - 2n_2}{n} \sum_{i;Y_i=0} X_i + \frac{n_0 - n_2}{n} \sum_{i;Y_i=1} X_i + \frac{2n_0 + n_1}{n} \sum_{i;Y_i=2} X_i,$$

$$U_{POM} = \frac{-n_1 - n_2}{n} \sum_{i;Y_i=0} X_i + \frac{n_0 - n_2}{n} \sum_{i;Y_i=1} X_i + \frac{n_0 + n_1}{n} \sum_{i;Y_i=2} X_i.$$

Comparing the two score vectors U_{GEE} and U_{POM}, they only differ slightly in their coefficients for genotype groups $Y_i = 0$ and $Y_i = 2$.

However we note that their null models are quite different: in GEE, the multiple phenotypes (X_i) are regressed on the covariates (Z_i) under the null, and $p \times l$ number of parameters are estimated; in POM, genotype (Y_i) is regressed on the covariates, thus only l parameters are to be estimated. For a large p (the dimension of X_i) and small n setting, fitting the GEE null model is likely to fail to converge; even if the GEE null model can be fitted, it becomes computationally more demanding as p grows. In contrast, fitting the POM null model does not suffer from these problems.

We also note that several authors have adopted a POM before for association testing with multiple traits: O'Reilly et al. [58] proposed the likelihood ratio test, while Wang et al. [57] derived the score test for POM. Both approaches assume a large samples size, which ensures a full-ranked estimate of the covariance matrix Σ. Compared to these approaches, the proposed method is useful for small n and large p settings. More importantly, even in the small p setting, our proposed adaptive test can outperform the classical likelihood ratio test and score test, as to be shown and demonstrated in other contexts [75, 28].

3.3 Results

3.3.1 Real data example

3.3.1.1 Testing with MRI phenotypes for $n > p$

The proposed POM-aSPU test was applied to an $n > p$ setting and empirically compared with the GEE-aSPU test [28]. We considered some candidate SNPs: rs429358, rs2075650, rs7526034, rs10932886, rs7647307, rs7610017, rs4692256 and rs6463843, which

were shown be strongly associated with some quantitative imaging traits [8]. From the ADNI-1 baseline scans, the cortical thicknesses for 68 ROIs were extracted based on the Desikan-Killany atlas [77]. The sample size was $n = 638$ with 145 ADs, 182 normal controls (CNs) and 311 subjects with minor cognitive impairment (MCIs).

We considered two different sets of multiple phenotypes. The first was a set of cortical thicknesses from all 68 ROIs ($p = 68$), and the second was a subset of only 12 ROIs related to the default mode network (DMN). DMN is a network of brain regions that are active when the individual is at wakeful rest, which includes left and right inferior parietal, inferior temporal, medial orbitofrontal, parahippocampal, precuneus and posterior cingulate [78, 72]. For covariates, gender, handedness and age measured at baseline were included. Permutation based POM-aSPU and GEE-aSPU tests were applied; the number of permutation was set at $B = 10^3$ at first, but was increased up to $B = 10^8$, if an obtained p-value was less than $5/B$.

Tables 3.1 and 3.2 report the p-values from the POM-aSPU and GEE-aSPU tests when the cortical thicknesses for DMN and all 68 regions were used as phenotypes. Both tests identified rs429458 to be associated with the cortical thicknesses in DMN, but not in all ROIs. APOE genotype (rs429358) is known to influence cortical thinning in Alzheimer's disease [79, 65, 80].

3.3.1.2 GWAS scan with rs-fMRI phenotypes for $n < p$

We conducted a GWAS scan for functional connectivity in the default mode network (DMN) with rs-fMRI data. We obtained rs-fMRI data and genotype data from ADNI-2. At each of 116 ROIs, neuronal activity was measured in BOLD time series. 18 ROIs related to DMN were defined as nodes. The selected nodes included left/right sides of superior frontal cortex, medial prefrontal cortex, ventral anterior cingulate cortex, posterior cingulate cortex parahipppocampal cortex, inferior parietal cortex, angular, middle temporal gyrus, and inferior temporal cortex [81, 72, 82]. Given a set of nodes, functional connectivity between every pair of 18 nodes was calculated with the Pearson's correlation of the two time series [83, 68, 76]. A total of $p = 18 \times (18 - 1)/2$ unique pairwise correlations was estimated, and Fisher's z-transformation was applied to each connection. For genotype data, we included all SNPs with a minor allele frequency (MAF) ≥ 0.05, genotyping rate more than 90%, and surviving the Hardy-Weinberg

equilibrium test with a p-value > 0.001. After all rounds of quality control, 578,175 SNPs remained. There were $n = 134$ subjects, consisting of 24 ADs, 22 late-onset MCIs (LMCIs), 44 early-onset MCIs (EMCIs), 20 subjects with symptoms of memory loss (SMCs) and 24 CNs.

The POM-aSPU test was applied to each of 578,175 SNPs to test its association with the DMN functional network after adjusting for age and gender. The p-values for the top three SNPs from the POM-aSPU test are reported in Table 3.3. The last column of Table 3.3 shows the $\hat{\gamma}$ values by which the SPU($\hat{\gamma}$) gave the minimum p-value among the SPU(γ) tests applied. SNP rs6663388 showed the strongest association with functional connectivity in DMN, but it was not in any gene. The second most significant SNP was rs11982066, in gene SEMA3E; this gene was selected for predicting survival/onset age of Parkinson disease [84], and also related to dysfunction in DMN [85]. Gene NRP1 (rs2804498) has been implicated in Alzheimer disease, combined with another gene SEMA3A [86]. Among the SPU(γ) tests applied with $\gamma \in \{1, ..., 8, \infty\}$, SPU(8) showed the minimum p-value for testing rs6663388, implying that one or few traits were associated with the SNP with relatively large effect sizes. SNPs rs11982066 and rs2804498 were given the minimum p-values by SPU(2) among the applied SPU(γ) tests, suggesting possibly many weak associations across the traits. Figures 3.1 is a LocusZoom plot [87] for the top three SNPs. Figures 3.2 and 3.3 illustrate a Q-Q plot and a Manhattan plot from the GWAS scan with the POM-aSPU test. All inflation factors were reasonable (with λ as shown in Q-Q plots), and the aSPU test succeeded in combining the significant associations identified by the individual SPU(γ) tests as presented in Manhattan plots. Although none of the SNPs was significant at the genome-wide significance level, it was perhaps due to the small sample size.

For testing associations with regularized networks, we applied the doubly adaptive test by thresholding the empirical correlation matrix with candidate thresholds $t \in \{0, 0.1, 0.2, ..., 0.9\}$. After thresholding, Fisher's z-transformation was applied to each connection. The p-values from the top four significant SNPs are presented in Table 3.4. SNPs rs1412096 in gene PTPRD and rs7276462 in gene GRIK1 were additionally identified, which were discussed as possible triggers to AD [88, 89, 90]. Figures 3.4 is a LocusZoom plot for the top four SNPs and Figure 3.5 illustrates a Q-Q plot and a Manhattan plot from the GWAS scan with the POM-daSPU test.

3.3.2 Simulations

3.3.2.1 Simulation set-ups

We carried out a small simulation study to investigate the performance of the proposed method as compared with the GEE-based tests [28] and the POM-based score test [57]. The first simulation set-up resembled an association pattern between SNP rs2075650 and DMN cortical thicknesses ($p = 12$) in Table 3.1. We assumed each phenotype to have possibly different inheritance modes, additive or dominant. Subjects were classified into 3 groups depending on the genotype score $Y_i \in \{0, 1, 2\}$. To sketch the simulation setup, we obtained the mean value of each individual phenotype for each genotype group. Let $\mu_j = (\mu_{j1}, ..., \mu_{jp})'$ be a vector for phenotype means for subject group $j \in \{0, 1, 2\}$. Figure 3.6(a) illustrates the mean cortical thicknesses of the 12 DMN regions for each genotype group as obtained from the ADNI-1 data. To mimic this pattern, we selected 7 traits (i.e. traits 1, 3, 4, 7, 8, 9, 12) to have an additive inheritance mode while the other 5 traits to have a dominant one. Figure 3.6(b) illustrates the mean values of individual phenotypes in simulated data, which resembles (a).

We defined the mean phenotype of j genotype group as

$$\mu_j = \beta_0 + \beta_j * j$$

with $\beta_j = (\beta_{j1}, ..., \beta_{jp})'$ for $j \in \{0, 1, 2\}$. To have both additive and dominant modes as in real data, we defined $\beta_{2k} = \beta_{1k}$ for an additive model, but set $\beta_{2k} = \beta_{1k}/2$ for a dominant mode. To mimic the real data, β_0, β_1 and the covariance matrix Θ of the multiple phenotypes were estimated after regressing out the genotype score of rs2075650 over the DMN cortical thickness measures.

The simulation procedure was the following. First, a genotype score (Y_i) was generated from a Bernoulli distribution, $Y_i \sim Ber(\text{MAF})$ with a given MAF. For set-up 1, MAF was defined at 0.1. Then multiple phenotypes $X_i = (X_{i1}, ..., X_{ip})'$ were simulated from a linear model:

$$X_i = \beta_0 + \phi\beta_1 \cdot I(Y_i = 1) + \phi\beta_2 \cdot I(Y_i = 2) + \epsilon_i.$$

where $\epsilon_i \sim \mathcal{MN}(0, \Theta)$. Here we introduced a scaling factor ϕ to control the association strength between X_i and Y_i. Under the null hypothesis of no association, $\phi = 0$; on the other hand, $\phi = 1$ gave the same association strength as that in the real data.

Similarly, the second simulation set-up was built on the association pattern between SNP rs429358 and the cortical thicknesses from all brain regions ($p = 68$) in Table 3.2. Among 68 phenotypes, we designated 59 phenotypes to have a dominant inheritance mode, while 5 to have an additive one, and the remaining 4 traits were always not associated with the SNP. For set-up 2, MAF was defined at 0.3.

By default, we considered a sample size $n = 1000$ at each simulated dataset. Empirical Type I error rates and power were evaluated based on 1000 replicates at significance level $\alpha = 0.05$ for each simulation scenario. For simulation- and permutation-based SPU and aSPU tests, we used $B = 1000$.

3.3.2.2 Type I error and power

Tables 3.5 and 3.6 report type I error rates ($\phi = 0$) and power ($\phi > 0$) for simulations. Type I error rates were well controlled by all methods applied. In simulation set-up 1 (Tables 3.5) where the majority of the multiple traits (7 out of 12) was linearly related to the genotype score (with an additive inheritance mode), GEE-based tests gave slightly higher power than the corresponding POM-based tests as expected. The SPU and aSPU tests showed better performance than the classical score test. In simulation set-up 2 (Tables 3.6) where 57 of the 68 phenotypes had a dominant inheritance mode, the POM-based tests had slightly higher power than the corresponding GEE-based tests. The score test performed better than the aSPU tests in both GEE and POM.

Depending on the simulation setting, either GEE-based tests or POM-based tests could slightly outperform the other, yet their overall performance was quite similar.

3.4 Conclusions

We have presented a new adaptive association test for multiple traits-single SNP association in a proportional odds model. From the analyses of the ADNI data and simulated data, we observed that the performance of the proposed POM-based tests was similar to that of the corresponding GEE-based tests (Table 3.5 and Table 3.6), as supported by an analysis of the similar score vectors from the two models. Nevertheless, the POM is more robust to the assumed inheritance mode, and more importantly, is computationally more efficient than GEE (Section 3.2.4). Moreover, the proposed POM-based tests

are applicable to high dimensional setting, for which functional connectivity phenotypes were employed as an example (Section 3.3.1.2). In the example, we observed that some, but not all, detected associations likely came from accumulating weak effects of individual traits (Table 3.3 and Table 3.4), and hence expect that our proposed POM-aSPU test is promising in identifying both joint weak associations and sparse strong signals, both of which may appear but are unknown in practice.

Table 3.1: P-values for association testing between DMN cortical thickness and each candidate SNP

			POM					GEE				
SNP	chr	maf	Score	SPU(1)	SPU(2)	SPU(∞)	aSPU	Score	SPU(1)	SPU(2)	SPU(∞)	aSPU
rs429358	19	0.30	5.17e-05	1.00e-08	1.00e-08	1.30e-04	2.00e-08	2.40e-07	1.90e-08	2.70e-08	5.00e-04	2.90e-08
rs2075650	19	0.25	1.35e-05	2.00e-07	1.00e-07	2.38e-03	9.00e-07	1.52e-06	1.90e-07	2.70e-07	1.32e-03	4.90e-07
rs7526034	1	0.12	2.53e-02	7.00e-04	6.00e-04	3.00e-03	1.30e-03	3.50e-02	3.00e-04	3.00e-04	1.80e-03	6.00e-04
rs10932886	2	0.32	2.89e-03	1.10e-03	9.00e-04	1.29e-02	2.00e-03	4.26e-03	2.80e-03	2.60e-03	1.44e-02	5.70e-03
rs7647307	3	0.44	7.34e-03	9.20e-03	7.80e-03	4.38e-02	1.52e-02	2.10e-03	8.00e-03	6.40e-03	2.20e-02	1.26e-02
rs7610017	3	0.04	5.65e-01	1.38e-01	1.93e-01	3.16e-01	2.16e-01	6.37e-01	1.63e-01	2.39e-01	4.02e-01	2.49e-01
rs4692256	4	0.46	8.42e-02	1.02e-01	6.88e-02	9.45e-02	1.04e-01	1.83e-01	8.24e-02	7.82e-02	1.11e-01	1.29e-01
rs6463843	7	0.47	1.04e-02	6.00e-04	8.00e-04	1.94e-02	1.20e-03	6.61e-03	3.00e-04	5.00e-04	1.93e-02	7.00e-04

Table 3.2: P-values for association testing between 68 regions' cortical thickness and each candidate SNP

			POM					GEE				
SNP	chr	maf	Score	SPU(1)	SPU(2)	SPU(∞)	aSPU	Score	SPU(1)	SPU(2)	SPU(∞)	aSPU
rs429358	19	0.30	6.15e-04	1.10e-06	7.00e-07	2.70e-06	3.60e-06	1.55e-06	1.50e-05	4.20e-06	6.46e-03	1.35e-05
rs2075650	19	0.25	4.99e-02	1.00e-05	1.00e-05	4.90e-04	3.00e-05	9.07e-03	4.00e-07	5.60e-06	1.98e-02	3.60e-06
rs7526034	1	0.12	5.11e-01	6.00e-04	2.00e-04	2.00e-04	2.00e-04	1.64e-03	1.00e-04	1.00e-04	1.60e-03	2.00e-04
rs10932886	2	0.32	1.73e-02	1.09e-02	1.09e-02	3.30e-02	2.17e-02	2.05e-02	1.38e-02	1.90e-02	3.49e-02	2.85e-02
rs7647307	3	0.44	1.07e-02	8.90e-03	7.30e-03	4.00e-02	1.48e-02	5.57e-02	5.10e-03	1.00e-02	3.89e-01	1.16e-02
rs7610017	3	0.04	5.51e-01	1.69e-01	2.64e-01	6.89e-01	2.67e-01	5.94e-01	1.73e-01	3.14e-01	6.87e-01	2.82e-01
rs4692256	4	0.46	1.29e-01	9.80e-02	3.81e-02	1.10e-02	2.11e-02	1.15e-01	7.24e-02	1.04e-01	1.65e-01	1.37e-01
rs6463843	7	0.47	3.85e-01	5.00e-04	5.00e-04	1.61e-02	1.10e-03	1.27e-01	1.00e-04	1.40e-03	1.67e-01	2.00e-04

Table 3.3: P-values of POM-aSPU test for functional connectivity in DMN

rs ID	chr	nearest gene	position	SPU(1)	SPU(2)	SPU(4)	SPU(∞)	aSPU	$\hat{\gamma}$
rs6663388	1	NA	165046596	1.10e-01	2.00e-05	2.22e-05	3.02e-07	6.70e-07	8
rs11982066	7	SEMA3E	82972395	1.16e-01	1.12e-06	1.34e-05	1.42e-05	9.96e-06	2
rs2804498	10	NRP1	33620707	1.90e-01	2.50e-06	1.79e-04	6.33e-03	1.23e-05	2

Table 3.4: P-values of POM-daSPU test for functional connectivity in DMN

rsID	chr	gene	position	daSPU	$(\hat{\gamma}, \hat{t})$
rs6663388	1	NA	165046596	2.90e-06	(2, 0.5)
rs1412096	9	PTPRD	9054913	9.92e-06	(2, 0)
rs2804498	10	NRP1	33620707	5.63e-06	(2, 0.2)
rs7276462	21	GRIK1	31429636	9.03e-07	(3, 0.3)

Table 3.5: Simulation setup 1: Type I errors (for $\phi = 0$) and power (for $\phi > 0$).

	POM									GEE								
		simulation-based				permutation-based					simulation-based					permutation-based			
ϕ	Score	SPU(1)	SPU(2)	SPU(∞)	aSPU	SPU(1)	SPU(2)	SPU(∞)	aSPU	Score	SPU(1)	SPU(2)	SPU(∞)	aSPU	SPU(1)	SPU(2)	SPU(∞)	aSPU
0	0.048	0.039	0.047	0.048	0.045	0.042	0.048	0.050	0.047	0.051	0.047	0.046	0.049	0.048	0.044	0.047	0.053	0.054
0.1	0.061	0.068	0.061	0.059	0.062	0.071	0.063	0.064	0.067	0.064	0.073	0.072	0.069	0.066	0.071	0.072	0.066	0.068
0.2	0.093	0.168	0.157	0.098	0.141	0.164	0.154	0.096	0.142	0.107	0.170	0.156	0.100	0.142	0.173	0.161	0.101	0.145
0.3	0.166	0.341	0.311	0.169	0.274	0.337	0.301	0.169	0.273	0.174	0.342	0.319	0.180	0.294	0.346	0.325	0.181	0.285
0.5	0.459	0.732	0.703	0.433	0.662	0.731	0.707	0.432	0.668	0.501	0.746	0.728	0.460	0.686	0.743	0.728	0.455	0.687
0.7	0.830	0.950	0.949	0.786	0.932	0.954	0.952	0.798	0.936	0.844	0.960	0.957	0.790	0.945	0.962	0.960	0.799	0.947
1.0	0.996	1.000	1.000	0.985	1.000	1.000	1.000	0.993	1.000	0.999	1.000	1.000	0.995	1.000	1.000	1.000	0.994	1.000

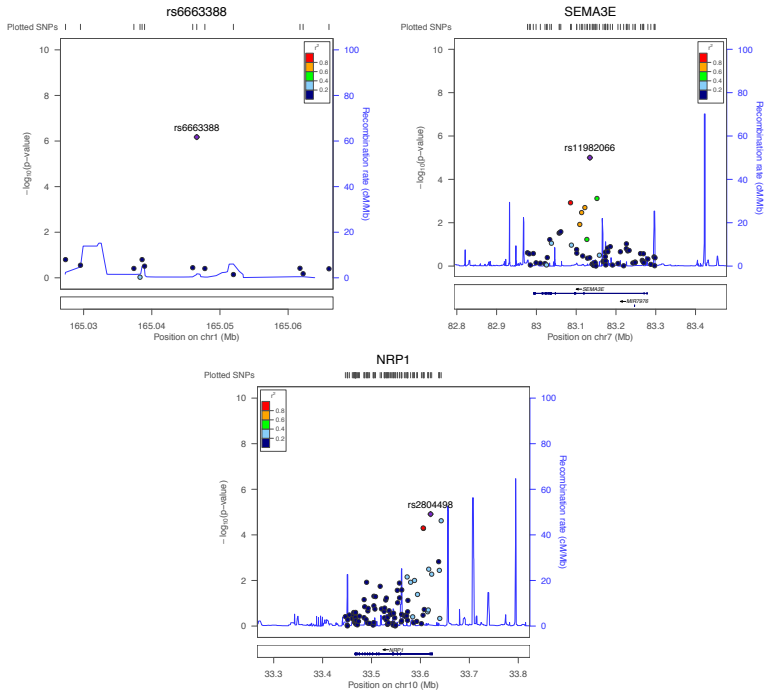


Figure 3.1: LocusZoom for top three SNPs for functional connectivity in DMN

Table 3.6: Simulation setup 2: Type I errors (for $\phi = 0$) and power (for $\phi > 0$).

	POM									GEE								
		simulation-based				permutation-based					simulation-based				permutation-based			
ϕ	Score	SPU(1)	SPU(2)	SPU(∞)	aSPU	SPU(1)	SPU(2)	SPU(∞)	aSPU	Score	SPU(1)	SPU(2)	SPU(∞)	aSPU	SPU(1)	SPU(2)	SPU(∞)	aSPU
0	0.047	0.049	0.052	0.057	0.050	0.051	0.051	0.057	0.049	0.041	0.044	0.049	0.056	0.050	0.051	0.048	0.053	0.050
0.1	0.074	0.089	0.065	0.068	0.083	0.091	0.062	0.064	0.084	0.071	0.087	0.066	0.068	0.075	0.083	0.068	0.065	0.081
0.2	0.221	0.227	0.136	0.114	0.172	0.222	0.131	0.113	0.170	0.187	0.197	0.122	0.111	0.157	0.196	0.130	0.114	0.171
0.3	0.616	0.428	0.292	0.184	0.345	0.426	0.283	0.186	0.339	0.542	0.391	0.262	0.183	0.317	0.389	0.263	0.189	0.317
0.4	0.931	0.667	0.520	0.326	0.578	0.664	0.527	0.321	0.578	0.893	0.616	0.476	0.327	0.538	0.611	0.479	0.328	0.528
0.5	0.998	0.849	0.740	0.480	0.789	0.848	0.740	0.479	0.789	0.994	0.801	0.695	0.471	0.747	0.807	0.693	0.474	0.740
0.7	1.000	0.984	0.973	0.777	0.966	0.986	0.971	0.773	0.970	1.000	0.974	0.956	0.768	0.956	0.974	0.956	0.768	0.958
1.0	1.000	1.000	1.000	0.976	1.000	1.000	1.000	0.974	1.000	1.000	1.000	1.000	0.978	1.000	1.000	1.000	0.977	1.000

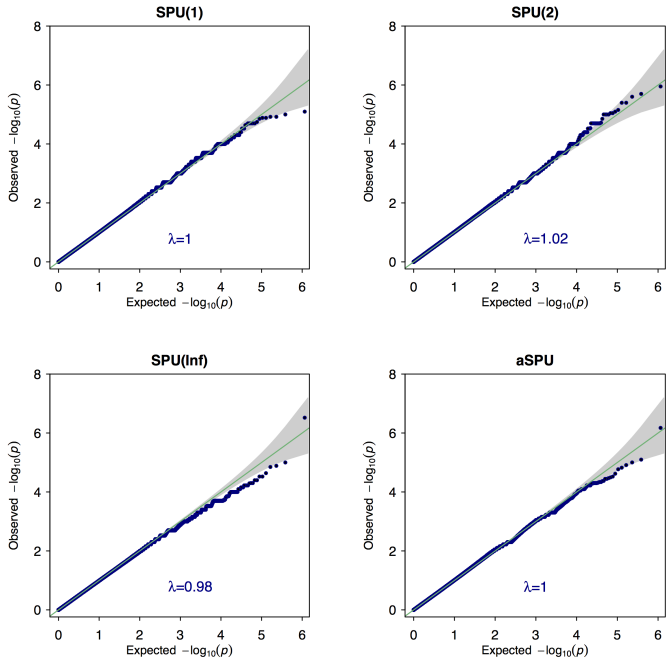


Figure 3.2: Q-Q plots from GWAS for function connectivity in DMN

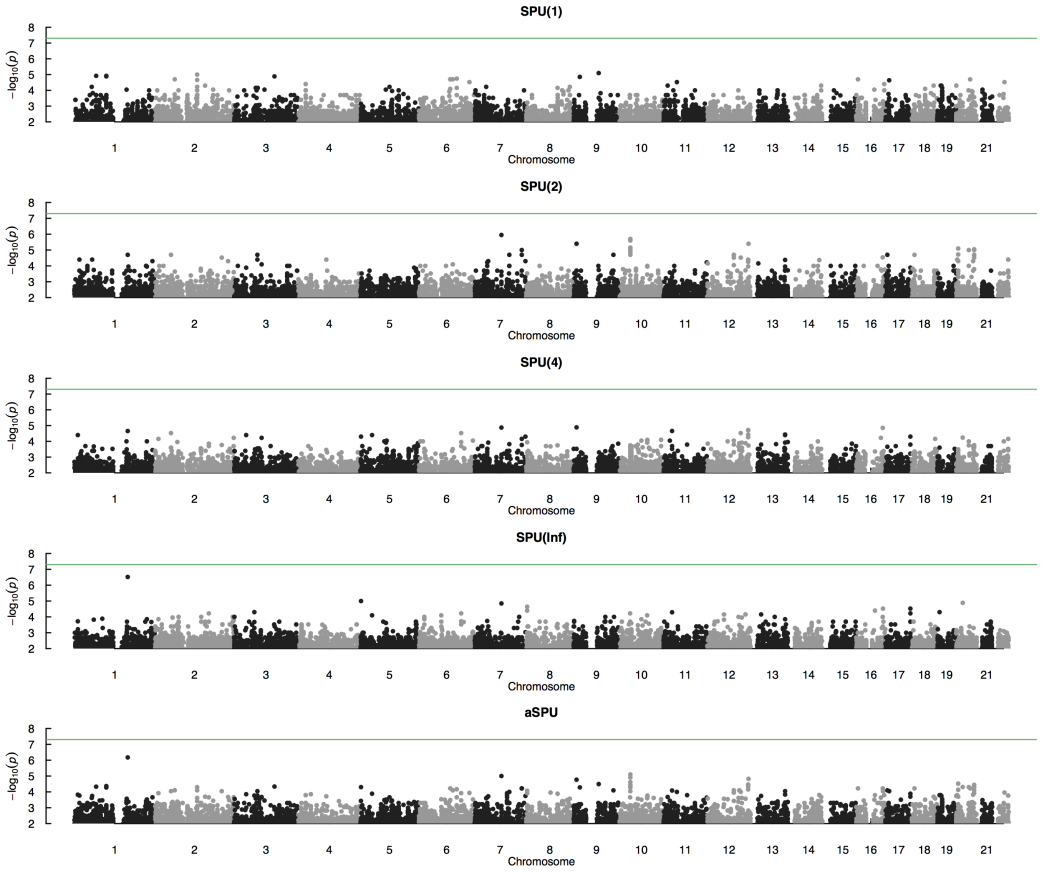


Figure 3.3: Manhattan plots from GWAS for function connectivity in DMN

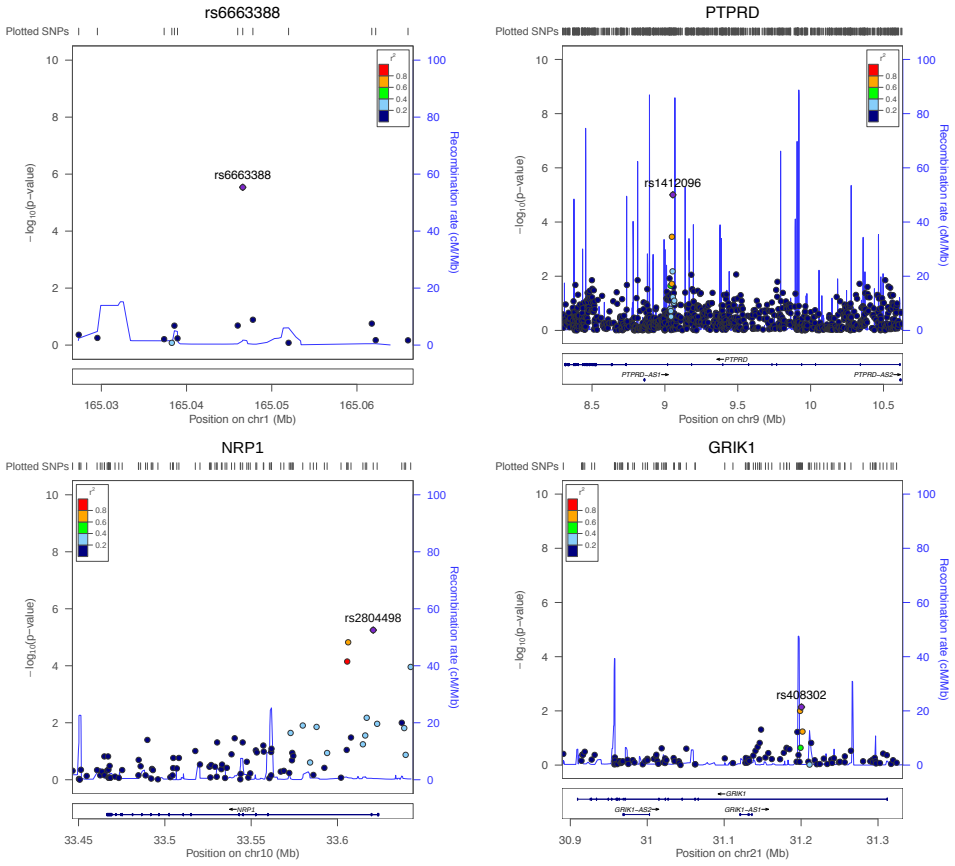


Figure 3.4: LocusZoom for top four SNPs for functional connectivity in DMN

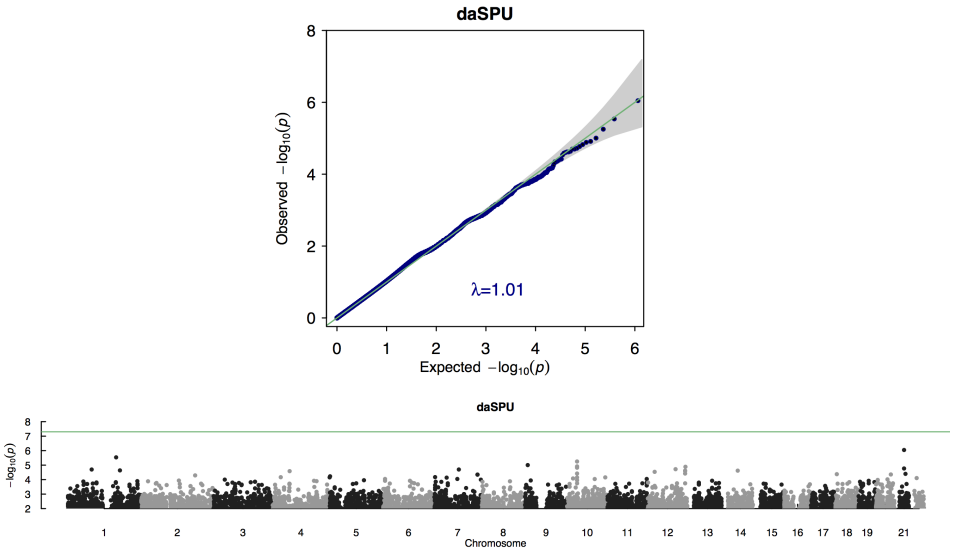
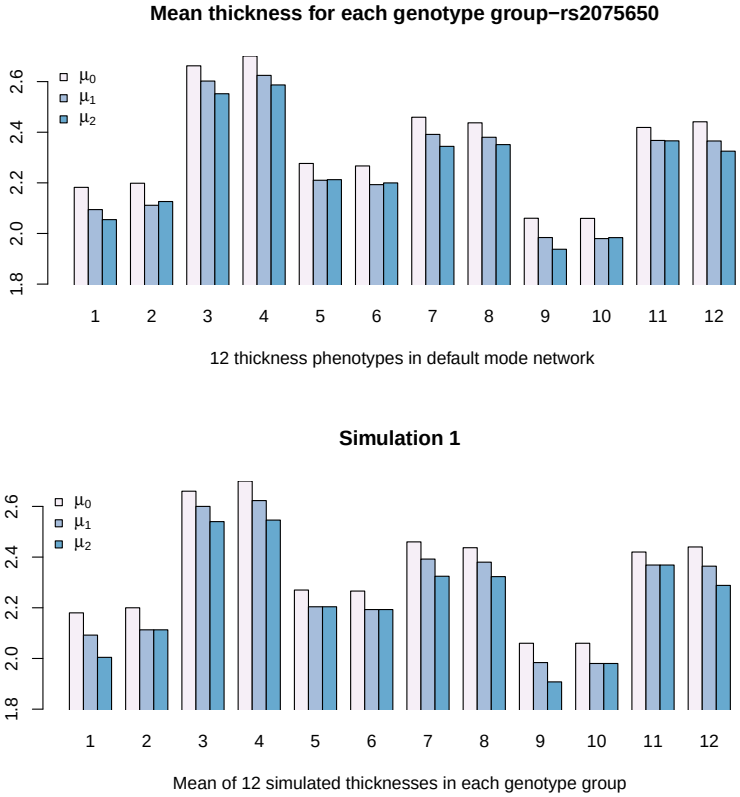


Figure 3.5: Q-Q plot and Manhattan plot from GWAS for sparse function connectivity in DMN

Figure 3.6: Mean phenotype in default mode network and simulation 1



Chapter 4

Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data

4.1 Introduction

Alzheimer's disease (AD) (MIM 104300) is the most common neurodegenerative disease, and every 67 seconds, someone in the U.S develops AD [91]. Currently there is no cure for AD, and most cases are diagnosed in the late stage of the disease. It is projected that the number of Americans age 65 years and older with AD will increase from 5.1 million in 2015 to 13.5 million in 2050, reflecting growth from an estimated 11% of the US senior population in 2015 to 16% in 2050, costing over $1.1 trillion in 2050 [92]. To advance our understanding of the initiation, progression and etiology of AD, Alzheimer's Disease Neuroimaging Initiative (ADNI) was started in 2004 and has continued since, collecting extensive clinical, genomic and multi-modal imaging data [2]. Many other genetic studies have been conducted, identifying multiple common and rare variants, shedding light on pathogenic mechanisms of AD [93, 94]. In particular, the APOEε4 allele has been consistently shown to be associated with AD. However, only

50% of AD patients carry an APOEε4 allele, suggesting the existence of other genetic variants contributing to risk for the disease [3]. 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 [4]. Hence, as for other common and complex diseases and traits, many more genetic factors underlying late onset AD are waiting to be discovered. One obvious but costly approach is to have a larger sample size. Alternatively, more powerful analysis methods are urgently needed. For example, in contrast to the popular single SNP-based analysis, novel gene- and pathway-based analyses may be more powerful in discovering additional causal variants. As demonstrated by Jones et al. [95], jointly analyzing functionally related SNPs sheds new light on the relatedness of immune regulation, energy metabolism and protein degradation to the etiology of AD. The reason is due to the well-known genetic heterogeneity and small effect sizes of individual common variants, as observed from published GWAS results [29]. To boost power in identifying aggregate effects of multiple SNPs, it may be promising to conduct association analysis at the SNP-set (or gene) level, rather than at the individual SNP level.

Another strategy is to use multiple endophenotypes, intermediate between genetics and the disease, for their potential to have stronger associations with genetic variants. In addition to boosting power, the use of intermediate phenotypes may provide important clues about causal pathways to the disease [12, 26]. A recent GWAS demonstrated the effectiveness of the strategy: some risk genes, such as FRMD6, were first identified to be associated with some neuroimaging intermediate phenotypes (e.g. hippocampal atrophy) [2], then were later validated to be associated with AD [96, 97]. A possibly useful but under-utilized intermediate phenotype is the brain default mode network (DMN), consisting of several brain regions of interest (ROIs) remaining active in the resting state. Brain activity in the DMN may explain the etiology of AD [98], and is a plausible indicator for incipient AD [62, 72, 99, 100, 101]. Since there is growing evidence that genetic factors play a role in aberrant default mode connectivity [64], it may be substantially more powerful to detect genetic variants associated with the DMN, a set of multiple intermediate phenotypes, than with AD.

Here we discuss gene-based multi-trait analysis, aiming at discovering genes associated with multiple traits such as the DMN. To date, several but not many methods have been proposed for gene-based multi-trait analysis [17, 16, 26, 102]. The simplest way is to use the minimum p-value (minP) test based on the most significant single SNP–single trait association, which however may lose power in the presence of multiple weak associations between multiple SNPs and multiple traits. Some methods, such as van der Sluis et al. [16] and M-TopQ25Stat [17], only utilize a few top association signals among the pairwise single SNP-single trait associations. Some methods based on principal components analysis (PCA) or principal components of heritability (PCH), originally proposed for multiple SNPs and a single trait [18, 19], may be also applied. However, these methods and canonical correlation analysis (CCA) [103] make use of only one or few top components, thus they share the same weakness of power loss in the presence of multiple associations; furthermore, the number of PCs may be difficult to determine [104]. Another extreme is the burden test [8, 17, 105], which is powerful in the presence of a dense association pattern, in which most SNP-trait pairs are associated with almost equal effect sizes and directions; otherwise, e.g. when the association directions of some SNP-trait pairs are different, it does not perform well (as is well known for analysis of rare variants). A compromise between the above two extremes is a variance-component test [26, 27], which is more robust to association density/sparsity and varying association directions. Nevertheless, as shown in the context for multiple rare variants and a single trait [75], it may still suffer from power loss in the presence of more sparse association patterns (i.e. when there are fewer associated SNP-trait pairs). A fundamental challenge in multivariate analysis is the lack of a uniformly most powerful test: any test may be powerful in some situations, but not in others. Nevertheless, we aim to construct an adaptive test such that it can maintain high power, not necessarily highest power, across a wide range of scenarios. In particular, the proposed test is adaptive at both the SNP and trait levels. Its key feature is the use of a weighting scheme to yield robust statistical power no matter whether the true and unknown association pattern is dense or sparse (or in whatever directions), and the weight is determined data-adaptively. In addition, some chosen weights correspond to several existing tests, including a burden test and a variance-component test. Therefore, the high power range of the proposed test covers those of the burden test and the variance-component test. Moreover, the

proposed test is based on the general framework of the generalized estimating equations (GEE), hence it is flexible with the capability to incorporate covariates and various types of traits [30]. It also avoids a difficulty in correctly specifying a joint multivariate distribution or likelihood for a set of multiple traits. Furthermore, we extend the proposed method to pathway analysis, in which it is adaptive to possibly varying gene-level associations.

We will compare the performance of the new test with several existing tests using both simulated and real data. The methods were applied to structural MRI data drawn from ADNI to identify genes associated with the DMN. In the GWAS, 277,527 SNPs were mapped to 17,557 genes, among which genes AMOTL1 on chromosome 11 and APOE on chromosome 19 were discovered by the new test to be significantly associated with the DMN. Notably, gene AMOTL1 was not detected by single SNP-based analyses. We also illustrate the application of the methods to the ADNI whole-genome sequencing (WGS) data, though no significant genes were identified, presumably due to a relatively small sample size.

In the following, we briefly review GEE and an existing method before introducing the new test in Methods. In Results, the new and several existing methods are compared with applications to the ADNI data and simulated data mimicking the ADNI data. We end with a short summary of the conclusions.

4.2 Methods

4.2.1 Review

4.2.1.1 Generalized estimating equations

Suppose for each individual $i = 1, ..., n$, we observe k traits $Y_i = (y_{i1}, ..., y_{ik})'$, q covariates $z_i = (z_{i1}, ..., z_{iq})'$ and a set of single nucleotide polymorphisms (SNPs) $x_i = (x_{i1}, ..., x_{ip})'$, with $x_{ij} \in \{0, 1, 2\}$. Denote $X_i = I \otimes x_i^{'}$ and $Z_i = I \otimes (1, z_i^{'})$, where I is a $k \times k$ identity matrix, and $\otimes$ represents the Kronecker product. We model the mean of the phenotypes $E(Y_i|X_i, Z_i) = \mu_i$, using a marginal generalized linear model

$$g(\mu_i) = Z_i\varphi + X_i\beta = H_i\theta \tag{4.1}$$

with $H_i = \begin{pmatrix} Z_i & X_i \end{pmatrix}$, parameters $\theta = (\varphi', \beta')'$, and a link function $g(.)$. The regression coefficients $\beta = (\beta_{11}, ..., \beta_{p1}, ..., \beta_{1k}, ..., \beta_{pk})'$ is a $pk \times 1$ vector, in which β_{jt} represents the effect of the jth SNP on the tth trait, while the element φ_{st} of $\varphi = (\varphi_{11}, ..., \varphi_{(q+1)1}, ..., \varphi_{1k}, ..., \varphi_{(q+1)k})'$ is the effect size of the sth covariate on the tth trait. Liang and Zeger [30] proposed estimating φ and β by solving GEE:

$$U_\theta = \sum_{i=1}^{n} D_i' V_i^{-1} (Y_i - \mu_i) = 0 \tag{4.2}$$

with $D_i = \partial \mu_i / \partial \theta'$ and $V_i = \phi A_i^{1/2} R_w(\alpha) A_i^{1/2}$, where ϕ is a dispersion parameter, $A_i = \text{diag}\{v(\mu_{i1}), ..., v(\mu_{ik})\}$ models the variances with a variance function $v(\mu_i)$, and $R_w(\alpha)$ is a working correlation matrix with possibly some unknown parameters α. Specifically, for quantitative traits (Y_i) with the identity link function (or more generally, for any generalized linear model with a canonical link function), the score vector U_θ and its variance-covariance matrix $Cov(U_\theta)$ are

$$U_\theta = (U_\varphi', U_\beta')' = \sum_{i=1}^{n} (Z_i \;\; X_i)' R_w^{-1} (Y_i - \mu_i),$$

$$Cov(U_\theta) = \sum_{i=1}^{n} (Z_i \;\; X_i)' R_w^{-1} (Y_i - \mu_i)(Y_i - \mu_i)' R_w^{-1} (Z_i \;\; X_i).$$

The covariance matrix can be partitioned according to the score components for φ and β: $Cov(U_\theta) = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$. For convenience, the working independence model is often used with R_w being as an identity matrix $I_{k \times k}$, as done in this chapter unless specified otherwise.

Our primary concern is to test for overall genetic effects with H_0: $\beta = 0$, while treating φ as nuisance parameters. To perform the score test, we evaluate the equation (4.1) under H_0. Under H_0, we have $g(\mu_i) = Z_i \varphi$, and the estimate of φ, denoted as $\widehat{\varphi}$, is the solution to the generalized score equation $U_{\varphi, \beta=0} = \sum_{i=1}^{n} Z_i' (Y_i - \mu_i) = 0$. The marginal mean is estimated by $\widehat{\mu}_i = g(Z_i \widehat{\varphi})^{-1}$.

For testing SNP-set effects, one considers the sub-components of the score vector for β:

$$U_\beta = \sum_{i=1}^{n} X_i' (Y_i - \widehat{\mu}_i). \tag{4.3}$$

U_β asymptotically follows a multivariate normal distribution $\mathcal{MN}(0, \tilde{\Sigma}_\beta)$ under H_0, where $\tilde{\Sigma}_\beta = V_{22} - V_{21}V_{11}^{-1}V_{12}$. U_β can be written as $U_\beta = (U_{11}, ..., U_{p1}, ..., U_{1k}, ..., U_{pk})'$. Each element U_{jt} measures the association strength between SNP j and trait t for $j = 1, ..., p$ and $t = 1, ..., k$, and is asymptotically proportional to β_{jt} in equation (4.1). $\beta_{jt} = 0$ implies there is no association between SNP j and trait t; similarly $U_{jt} = 0$ (or small) indicates no (or weak) association between SNP j and trait t.

For testing H_0, the GEE-Score test statistic is defined by

$$\text{GEE-Score} = U_\beta^{'}\tilde{\Sigma}_\beta^{-1}U_\beta.$$

Under H_0, the GEE-Score statistic asymptotically follows a central chi-squared distribution with pk degrees of freedom. When pk is large, this standard score test loses power for large degrees of freedom. Another way to draw inference, especially convenient when combining the score test with other tests as to be discussed later, is to simulate $U_\beta^{(b)} \sim \mathcal{MN}(0, \tilde{\Sigma}_\beta)$ for $b = 1, ..., B$ and obtain the null statistics $\text{GEE-Score}^{(b)} = U_\beta^{(b)'}\tilde{\Sigma}_\beta^{-1}U_\beta^{(b)}$. The p-value can be calculated as $P_{\text{Score}} = \sum_{b=1}^{B} I(\text{GEE-Score} \leq \text{GEE-Score}^{(b)})/(B+1)$, where $I(\cdot)$ denotes the indicator function.

For ease of notation, we suppress β and take $U = U_\beta$ and $V = \tilde{\Sigma}_\beta$ hereafter.

4.2.1.2 An adaptive association test for a single SNP

Zhang et al. [28] proposed a class of sum of powered score (SPU) tests for testing association between an individual SNP and multiple traits, along with its data-adaptive version (aSPU). The SPU tests are a family of association tests based on the (generalized) score vector in the GEE framework, aiming for at least one of them to be powerful in any given situation. With only a single SNP j, then the score vector reduces to $U = (U_{j1}, ..., U_{jk})'$. The association between the SNP and k traits can be quantified with a test statistic

$$\text{SPU}(\gamma) = \sum_{t=1}^{k}(U_{jt})^\gamma$$

where a candidate integer $\gamma \geq 1$ is to be chosen from a pre-selected parameter set Γ; e.g. $\Gamma = \{1, 2, ..., 8, \infty\}$. The statistical power of an SPU(γ) test depends on the choice of $\gamma \in \Gamma$. When γ is an odd integer, the SPU(γ) test sums up the association

signals across all the traits, retaining high power if all or most of the multiple traits have an almost equal effect size in the same association direction. A special case is $\gamma = 1$, giving a burden test commonly used for rare variants. With an even γ, the SPU(γ) test will be more powerful when some traits have different association directions. In particular, the SPU(2) test is the same as the sum of squared score (SSU) test [106], closely related to multivariate distance matrix regression (MDMR) [107], kernel machine regression (KMR) [74] and variance-component tests [108]. Furthermore, as γ increases, the SPU test upweights the more strongly associated traits, while reducing the weights on other ones. In particular, when $\gamma \to \infty$ (as an even integer), only the maximum component of the score vector is used and the test statistic is defined as $\text{SPU}(\infty) = \max_{t=1}^{k} |U_{jt}|$. The SPU($\infty$) test is similar to the UminP test (when the variances of the score components are almost equal). To compute the significance of an SPU test, Monte Carlo (MC) simulations (or alternatively, permutations) are used; for $b = 1, ..., B$, the null score $U^{(b)} = (U_{j1}^{(b)}, ..., U_{jk}^{(b)})'$ is generated from $\mathcal{MN}(0, V)$, from which the null statistics $\text{SPU}(\gamma)^{(b)} = \sum_{t=1}^{k} (U_{jt}^{(b)})^{\gamma}$ can be obtained for each γ. Then the p-value can be calculated as $p_\gamma = [\sum_{b=1}^{B} I(\text{SPU}(\gamma) \leq \text{SPU}(\gamma)^{(b)}) + 1]/(B + 1)$.

However, it is not clear how to choose an optimal γ a priori for given data. Hence, Zhang et al [28] proposed an adaptive SPU (aSPU) test to extract association evidence from multiple SPU(γ) tests. The statistic of the aSPU test is the minimum p-value of SPU(γ)'s for some candidate values of γ:

$$\text{aSPU} = \min_{\gamma \in \Gamma} p_\gamma,$$

where p_γ is p-value of SPU(γ). By MC simulations (or permutations), the p-value of aSPU, along with those of all SPU(γ) tests, can be efficiently calculated based on the same set of the null statistics in a single layer.

4.2.1.3 Existing gene-based tests

We will compare the proposed test with several existing gene-based tests for multiple traits, including multivariate analysis of variance (MANOVA), MDMR with the Euclidean distance [107], multivariate KMR under linear kernel [26] and a multivariate functional linear model (MFLM) [102]. We would note that KMR can be derived

based on a random-effects model while MFLM is built on a fixed effect model. For implementation, R package `vegan` was used for MDMR; R code for KMR and MFLM was downloaded from the authors' websites, `http://www4.stat.ncsu.edu/~maity/software.html` and `https://www.nichd.nih.gov/about/org/diphr/bbb/software/fan/Pages/default.aspx` respectively. Since KMR [26] was computationally slow, and excluded from the simulation studies.

4.2.2 New Methods

4.2.2.1 An adaptive test

We introduce a novel gene-based adaptive sum of powered score test for a set of multiple traits, denoted as *aSPUset*, by extending the single SNP-based test of Zhang et al. [28]. Suppose that there are p SNPs in a gene and k traits of interests. Recall that $U = (U_{11}, ..., U_{p1}, ..., U_{1k}, ..., U_{pk})'$ is the generalized score vector of length pk in GEE, and V is the $pk \times pk$ covariance matrix of the score vector; each element of the score, U_{jt} quantifies the association between SNP j and trait t. In practice, the true and unknown association patterns across multiple SNPs and multiple traits are complex: some SNPs may be associated with some traits, but not with other traits; different SNPs may be associated with different subsets of the traits with varying association strengths and directions. Since the inclusion of non-associated SNPs and traits in a test statistic could reduce the power of the test, we may want to give higher weights to more likely associated SNPs and traits. However, how much to optimally overweight these likely associated SNPs and traits depends on the true association pattern, which is unknown. The aSPUset test employs two positive integer parameters, γ_1 and γ_2, to control the degrees of weighting over the SNPs and over the traits respectively, and the two parameters are chosen data-adaptively. A larger γ_1 puts more weights on the SNPs more likely to be associated with a given trait, while a larger γ_2 upweights the traits more strongly associated with the SNPs.

We build the test statistic as follows. For each trait t, $S(\gamma_1;\ t)$ quantifies the association between the single trait and multiple SNPs, then SPU(γ_1, γ_2) combines the

single trait-based statistics:

$$S(\gamma_1;\ t) = \big(\sum_{j=1}^{p}(U_{jt})^{\gamma_1}\big)^{1/\gamma_1}, \qquad \text{SPU}(\gamma_1,\gamma_2) = \sum_{t=1}^{k}\big(S(\gamma_1;\ t)\big)^{\gamma_2}. \tag{4.4}$$

Here candidate integers $\gamma_1 \geq 1$ and $\gamma_2 \geq 1$ are to be chosen from two pre-selected parameter sets Γ_1 and Γ_2. We used $\Gamma_1 = \Gamma_2 = \{1, 2, ..., 8, \infty\}$, due to the good performance in our numerical studies.

In $S(\gamma_1;\ t)$, $(U_{jt})^{\gamma_1}$ can be re-written in an alternative form $(U_{jt})^{\gamma_1} = U_{jt}^{\gamma_1-1}U_{jt} = w_{jt}U_{jt}$. $w_{jt} = U_{jt}^{\gamma_1-1}$ is a weight for each score element, which reflects the association strength (and direction) between SNP j and trait t of the given data. With $\gamma_1 = 1$, the SPU test weights each SNP equally, and yields the highest power if all the SNPs are associated with the trait t with similar effect sizes and association direction (i.e. all positive or all negative). When the subset of SNPs are associated with the trait t, or their association directions are different, $\text{SPU}(\gamma_1 = 2, \gamma_2)$ is often more powerful. As γ_1 increases, $\text{SPU}(\gamma_1, \gamma_2)$ puts heavier weights on the SNPs which are more strongly associated with the trait t. At the end, as the parameter approaches to ∞ (as an even integer), it only considers the most significant SNP, i.e. $\text{SPU}(\gamma_1 = \infty, \gamma_2) = \sum_{t=1}^{k}\big(\max_{j=1}^{p}|U_{jt}|\big)^{\gamma_2}$.

Similarly, γ_2 controls how much to up-weight the traits that are more likely to be associated with SNPs. $\text{SPU}(\gamma_1, \gamma_2 = 1)$ weights all traits equally and performs best when each trait is equally associated with the SNPs. Similarly, as γ_2 increases, the SPU test over-weights larger trait-based statistics $S(.;\ t)$; in an extreme case, as $\gamma_2 \to \infty$, we define $\text{SPU}(\gamma_1, \gamma_2 = \infty) = \max_{t=1}^{k}\big|S(\gamma_1;\ t)\big|$. If one is more interested in the most significantly associated single SNP-single trait pair, $\text{SPU}(\gamma_1 = \infty, \gamma_2 = \infty) = \max_{j,t}|U_{jt}|$ can be considered. Using various combinations of γ_1 and γ_2, one can target and fit different association patterns across multiple SNPs and multiple traits, including their varying sparsity levels. As a result, the $\text{SPU}(\gamma_1, \gamma_2)$ tests cover several existing tests as special cases as to be shown.

The aSPUset test chooses (γ_1, γ_2) data-adaptively by taking the minimum p-value of $\text{SPU}(\gamma_1, \gamma_2)$'s as the test statistic for candidates $\gamma_1 \in \Gamma_1$ and $\gamma_2 \in \Gamma_2$,

$$\text{aSPUset} = \min_{\gamma_1,\gamma_2} p_{\gamma_1,\gamma_2}.$$

To assess the significance of all the SPU(γ_1, γ_2) and aSPUset test, we use either permutations or MC simulations in a single layer to obtain their p-values. The permutation-based method is useful when the covariance matrix (V) is not easy to estimate (e.g. in a high dimensional setting) or when the usual Normal asymptotics may not hold (e.g. n is not large compared to pk); in contrast, the simulation-based method is more restrictive but computationally more efficient. For the permutation-based method, residual terms $\text{res}_i = Y_i - \widehat{\mu}_i$ in equation (4.3) are permuted to generate $\text{res}_i^{(b)}$ for $b = 1, ...B$, from which the null score vector $U^{(b)}$ is computed as $U^{(b)} = \sum_{i=1}^{n} X_i' \text{res}_i^{(b)}$. Alternatively, for the simulation method, we simulate the null score vectors independently from the null distribution: $U^{(b)} \sim \mathcal{MN}(0, V)$ for $b = 1, ...B$.

In either case, the null statistics SPU$(\gamma_1, \gamma_2)^{(b)}$ can be calculated from the null score vectors $U^{(b)}$ for $b = 1, ..., B$. Because all SPU(γ_1, γ_2) tests are based on the same null score vectors $U^{(b)}$, we just need to simulate one set of null scores and efficiently compute the null statistics, SPU$(\gamma_1, \gamma_2)^{(b)}$ tests simultaneously for candidate γ_1, γ_2's. Then the p-value of SPU(γ_1, γ_2) is

$$p_{\gamma_1,\gamma_2} = \frac{1 + \sum_{b=1}^{B} (I(|\text{SPU}(\gamma_1, \gamma_2)^{(b)}| \geq |\text{SPU}(\gamma_1, \gamma_2)|)}{B + 1}.$$

We can also simultaneously and efficiently compute the p-value of the aSPUset test based on the same set of the null statistics being used for the SPU tests. Note that for each SPU$(\gamma_1, \gamma_2)^{(b)}$, we can calculate its p-value as
$p_{\gamma_1,\gamma_2}^{(b)} = [\sum_{l \neq b} (I(|\text{SPU}(\gamma_1, \gamma_2)^{l}| \geq |\text{SPU}(\gamma_1, \gamma_2)^{(b)}|) + 1]/B$. Denote its minimum as $p^{(b)} = \min_{\gamma_1,\gamma_2} p_{\gamma_1,\gamma_2}^{(b)}$. Then the significance of aSPUset test is obtained as

$$P_{\text{aSPUset}} = \frac{\sum_{b=1}^{B} I(|p^{(b)}| \leq |\text{aSPUset}|) + 1}{B + 1}.$$

4.2.2.2 Extensions

As shown by Zhang et al. [28], in some but not all situations, the GEE-Score test may perform better than the aSPU test for a single SNP and multiple traits; the opposite is true too. Hence, to take advantage of both tests, we combine them by taking their minimum p-value to form a new test statistic,

$$\text{aSPUset-Score} = \min\left(P_{\text{aSPUset}}, P_{\text{Score}}\right). \tag{4.5}$$

Its p-value can be calculated using simulations or permutations as for aSPUset. The null statistic GEE-Score$^{(b)}$ is obtained from the same score $U^{(b)}$ which is used for SPU$(\gamma_1, \gamma_2)^{(b)}$. Hence the null statistics for SPU$(\gamma_1, \gamma_2)^{(b)}$ and GEE-Score$^{(b)}$ can be computed simultaneously.

We can also consider a variance-weighted version of the SPU and aSPUset tests, called the SPUw and aSPUw-set respectively. Each diagonal element of covariance matrix (V) corresponds to the variance of the individual score element U_{jt}; denote the variance of U_{jt} as V_{jt}. The SPUw test is defined with statistic

$$\text{SPUw}(\gamma_1, \gamma_2) = \sum_{t=1}^{k} \left\{ \left[\sum_{j=1}^{p} (U_{jt}/\sqrt{V_{jt}})^{\gamma_1} \right]^{1/\gamma_1} \right\}^{\gamma_2}.$$

The aSPUw-set test statistic is defined as the one taking the minimum p-value of the multiple SPUw(γ_1, γ_2) tests in the same way as that for aSPUset and SPU(γ_1, γ_2). The SPUw and aSPUw-set tests are invariant to the scale of each trait, and hence may be useful when it is unclear how to standardize multiple traits that are in different scales. However, standardizing the traits (such that their sample variances are all equal to one) may or may not be beneficial; often, the power of the unweighted SPU tests and that of the weighted ones are similar as shown before in other contexts [75, 28].

4.2.2.3 Relationships with other methods

The SPU tests are closely related to some existing tests, covering some as special cases. We consider the case without covariates, since several methods are only applicable to the case without covariates.

Without loss of generality we center both $Y_i = (y_{i1}, y_{i2}, ..., y_{ik})'$ and $x_i = (x_{i1}, x_{i2}, ..., x_{ip})'$ to have their sample means $\sum_{i=1}^{n} Y_i/n = 0$ and $\sum_{i=1}^{n} x_i/n = 0$, and rewrite data in a design matrix. Denote Λ as an $n \times p$ matrix such that each row contains subject i's genotype $x_i = (x_{i1}, ..., x_{ip})'$ and Θ as an $n \times k$ matrix such that each row consists of subjects' multiple traits $Y_i = (y_{i1}, ..., y_{ik})'$. Multivariate analysis can be derived by partitioning of the total sum of squares and cross products (SSCP) matrix, the inner product $\Theta'\Theta$. According to the multivariate linear model, $\Theta = \Lambda \mathrm{B} + \mathrm{E}$, where B is the matrix of model parameters, E is the matrix of errors, the fitted value matrix is defined

as $\widehat{\Theta} = \Lambda\widehat{\mathrm{B}} = \Lambda(\Lambda^{'}\Lambda)^{-1}\Lambda^{'}\Theta = \mathrm{H}\Theta$ and the matrix of residuals is R= $\Theta - \widehat{\Theta} = (\mathrm{I} - \mathrm{H})\Theta$, where H is a hat matrix.

We define each covariance estimate as follows. $S_x = \frac{1}{n}\Lambda'\Lambda$ is a $p \times p$ covariance estimate for genotype scores $x_i = (x_{i1}, ..., x_{ip})'$, and $S_y = \frac{1}{n}\Theta'\Theta$ is a $k \times k$ covariance estimate among k multiple traits $Y_i = (y_{i1}, ..., y_{ik})'$. The covariance estimates between two sets of variable x_i and Y_i are defined as $S_{yx} = \frac{1}{n}\Theta'\Lambda$ and $S_{xy} = \frac{1}{n}\Lambda'\Theta$. Note that tr(A) stands for sum of diagonal elements of a matrix A; vec(A) represents a linear transformation which converts the matrix (A) into a column vector.

SPUw(2,2) and M-MeanStat; SPUw(∞,1) and M-Max Guo et al. [17] proposed a set of nonparametric methods for gene-based multiple trait association analysis, called M-MeanStat, M-MaxStat, and M-TopQ25Stat. Each method of Guo et al. [17] is built on a generalized Kendall's tau (τ), which quantifies the pairwise association between a single SNP and a single trait. Between SNP j and trait t, the pairwise association is defined by $\tau_{jt} = \sum_{i=1}^{n} x_{ij}(y_{it} - \bar{y}_t) = \sum_{1}^{n} x_{ij}y_{it}$, which follows a normal distribution asymptotically with mean zero and variance $\mathrm{var}(\tau_{jt}|y_t) = \sum_{i=1}^{n} \mathrm{var}(x_{ij})y_{it}^2$ under the null hypothesis. Guo et al. [17] defined the generalized Kendall's tau statistic, $T_{jt} = \tau_{jt}^2 \mathrm{var}(\tau_{jt}|y_t)^{-1} \sim \chi_1^2$. Based on this, Guo et al. [17] proposed M-MeanStat and M-MaxStat;

$$\begin{aligned}
\text{M-MeanStat} &= \frac{1}{p}\sum_{t=1}^{k}\sum_{j=1}^{p} T_{jt} \propto \sum_{t=1}^{k}\sum_{j=1}^{p} \frac{(\sum_{i=1}^{n} x_{ij}y_{it})^2}{\sum_{i=1}^{n} \mathrm{var}(x_{ij})y_{it}^2} \approx \sum_{t=1}^{k}\sum_{j=1}^{p} \left(\frac{\sum_{i=1}^{n} x_{ij}y_{it}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2 y_{it}^2}} \right)^2, \\
\text{M-MaxStat} &= \sum_{t=1}^{k} \max_{j=1}^{p} T_{jt} = \sum_{t=1}^{k} \max_{j=1}^{p} \frac{(\sum_{i=1}^{n} x_{ij}y_{it})^2}{\sum_{i=1}^{n} \mathrm{var}(x_{ij})y_{it}^2} \approx \sum_{t=1}^{k} \max_{j=1}^{p} \left(\frac{\sum_{i=1}^{n} x_{ij}y_{it}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2 y_{it}^2}} \right)^2.
\end{aligned} \tag{4.6}$$

If a canonical link function and a working independence model are used in GEE, the

test statistics of SPUw$(2,2)$ and SPUw$(\infty,1)$ are defined by

$$\begin{aligned}
\text{SPUw}(2,2) &\propto \sum_{t=1}^{k}\sum_{j=1}^{p}\left(\frac{\sum_{i=1}^{n} x_{ij}y_{it}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2 \text{var}(y_{it})}}\right)^2 \approx \sum_{t=1}^{k}\sum_{j=1}^{p}\left(\frac{\sum_{i=1}^{n} x_{ij}y_{it}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2 y_{it}^2}}\right)^2, \\
\text{SPUw}(\infty,1) &\propto \sum_{t=1}^{k}\max_{j=1}^{p}\left|\frac{\sum_{i=1}^{n} x_{ij}y_{it}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2 \text{var}(y_{it})}}\right| \approx \sum_{t=1}^{k}\max_{j=1}^{p}\left(\frac{\sum_{i=1}^{n} x_{ij}y_{it}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2 y_{it}^2}}\right)^2. \qquad (4.7)
\end{aligned}$$

Comparing the two sets of statistics in equations (4.6) and (4.7), we see that M-MeanStat and SPUw$(2,2)$, and M-Max and SPUw$(\infty,1)$ are approximately equivalent respectively.

SPU(2,2) and MDMR The SPU(2,2) test has connections to several other tests. Zhang et al. [28] showed that when testing on a single SNP, the SPU(2,2) test under the GEE working independence model is equivalent to MDMR with the Euclidean distance. However, for testing multiple SNPs, the equivalence does not hold. Under the working independence model, the test statistic of SPU(2,2) is stated as

$$\text{SPU}(2,2) = \sum_{t=1}^{k}\sum_{j=1}^{p}\big(\sum_{i=1}^{n} x_{ij}y_{it}\big)^2 = \text{tr}\big(\Lambda'\Theta\Theta'\Lambda\big). \qquad (4.8)$$

MDMR is a nonparametric modification of traditional Fisher's MANOVA [107]. Wessel and Schork [109] and Zapala and Schork [110] introduced the method to applications in genetics and genomics. For single trait, it is closely related to kernel methods [106, 111]. Suppose d_{ij} represents the distance between subject i and j; let $A = (a_{ij}) = (-1/2\ d_{ij}^2)$ and G its centered version. An F-statistic can be constructed to test the hypothesis that the p regressor variables have no relationship to variation in the distance or dissimilarity of the n subjects reflected in the $n \times n$ distance/dissimilarity matrix. The psuedo F-statistics of MDMR is defined by

$$F = \frac{\text{tr(HGH)}}{\text{tr(I}-\text{H)G(I}-\text{H)}}$$

If the Euclidean distance (i.e. L_2-norm) is used to construct the distance matrix (i.e. G$=\Theta\Theta'$), the MDMR test statistic is defined as

$$\text{MDMR} \propto \frac{\text{tr(H}\Theta\Theta'\text{H)}}{\text{tr(I}-\text{H)}\Theta\Theta'\text{(I}-\text{H)}} \propto \frac{1}{\text{tr}\big(\text{R}'\text{R}\big)/\text{tr}\big(\widehat{\Theta}'\widehat{\Theta}\big)} \propto \frac{1}{[\text{tr}\big(\widehat{\Theta}'\widehat{\Theta}\big)+\text{tr}\big(\text{R}'\text{R}\big)]/\text{tr}\big(\widehat{\Theta}'\widehat{\Theta}\big)} = \frac{\text{tr}\big(\widehat{\Theta}'\widehat{\Theta}\big)}{\text{tr}\big(\Theta'\Theta\big)}.$$

As usual, permutations are used to calculate p-values. Then $\mathrm{tr}\big(\Theta^{'}\Theta\big)$ is invariant across all permutations and can be ignored [106]. The test statistic arrives at

$$\mathrm{MDMR} \propto \mathrm{tr}\big(\widehat{\Theta}^{'}\widehat{\Theta}\big) = \mathrm{tr}\big(\Theta^{'}\Lambda(\Lambda^{'}\Lambda)^{-1}\Lambda^{'}\Theta\big) = \mathrm{tr}((\Lambda^{'}\Lambda)^{-1}\Lambda^{'}\Theta\Theta^{'}\Lambda\big). \tag{4.9}$$

If we have a single SNP to be tested, i.e. Λ is an $n \times 1$ matrix; the test statistic (4.9) reduces to $\mathrm{MDMR} \propto m^{-1}\mathrm{tr}(\Lambda^{'}\Theta\Theta^{'}\Lambda\big) \propto \mathrm{tr}(\Lambda^{'}\Theta\Theta^{'}\Lambda\big)$ with $\Lambda^{'}\Lambda = m$. Hence, SPU(2,2) and MDMR are equivalent for a single SNP and multiple traits, as established by Zhang et al [28]. However, for multiple SNPs and multiple traits, by comparing equations (4.8) and (4.9), we see that in general they are not equivalent.

SPU(2,2) and KMR KMR with the linear kernel is equivalent to SPU(2,2) if the working correlation matrix R_w of the latter in GEE is correctly specified as the true correlation matrix of Y_i.

With a working correlation matrix R_w in GEE, the SPU(2,2) test can be rewritten as

$$\mathrm{SPU}(2,2) = \mathrm{tr}\big(\Lambda^{'}\Theta R_w^{-1}R_w^{-1}\Theta^{'}\Lambda\big) = \mathrm{tr}(R_w^{-1}\Theta^{'}\Lambda\Lambda^{'}\Theta R_w^{-1}). \tag{4.10}$$

Maity et al. [26] introduced a multivariate phenotype association analysis by SNP set- or gene-based KMR. The authors assumed that the phenotypes are correlated while the individuals are independent. Suppose $\Psi = (\psi_{pq})$ is the true correlation matrix for k traits with $p = 1, ..., k$, and $q = 1, ..., k$. Define $V_0 = \Psi \otimes I_{n\times n}$, and a kernel matrix $\mathcal{K}_{nk\times nk}$. The score statistic under the null for KMR [26] is defined by

$$\mathrm{KMR} = \mathrm{vec}(\Theta)^{'}V_0^{-1}\ \mathcal{K}\ V_0^{-1}\mathrm{vec}(\Theta) = \mathrm{vec}(\Theta)^{'}V_0^{-1}\ \mathrm{diag}(K_1, ..., K_k)\ V_0^{-1}\mathrm{vec}(\Theta)$$

where each $K_1, ..., K_k$ is an $n \times n$ kernel matrix for each trait. Applying a linear kernel $K_1 =, ..., = K_k = \Lambda\Lambda^{'}$ yields

$$\begin{aligned}\mathrm{KMR} &= \mathrm{vec}(\Theta)^{'}V_0^{-1}\big(I_{k\times k} \otimes \Lambda\Lambda^{'}\big)V_0^{-1}\mathrm{vec}(\Theta) = \mathrm{vec}(\Theta\Psi^{-1})^{'}\big(I \otimes \Lambda\Lambda^{'}\big)\mathrm{vec}(\Theta\Psi^{-1}) \\ &= \mathrm{vec}(\Theta\Psi^{-1})^{'}\mathrm{vec}(\Lambda\Lambda^{'}\Theta\Psi^{-1}) = \mathrm{tr}(\Psi^{-1}\Theta^{'}\Lambda\Lambda^{'}\Theta\Psi^{-1}).\end{aligned} \tag{4.11}$$

KMR (equation 4.11) has the same test statistic as the GEE-SPU(2) test (equation 4.10) if the working correlation R_w is the true correlation structure of Y_i (i.e. $\Psi = R_w = Corr(Y_i|H_0)$). This illustrates the flexibility of our proposed test under GEE, in

contrast to the stronger modeling assumption in KMR. It is obvious that the SPU(1,1) test is a burden test, which is optimal if its implicit assumption that each SNP-trait pair is equally associated (with the same association direction) holds. Since KMR can be derived based on a random-effects model while the burden test is formulated based on a fixed-effects model, our proposed method can be regarded as combining results from both fixed- and random-effects models.

GEE-Score test and MANOVA As to be shown in our numerical studies, the GEE-Score test and MANOVA performed similarly; we establish the equivalence between the GEE-Score test and MANOVA with the Pillai-Bartlett trace. The GEE-Score test statistic with a working independence model in GEE is

$$\begin{aligned}\text{GEE-Score} &= \text{vec}(\Lambda^{'}\Theta)^{'}\big(S_y \otimes nS_x\big)^{-1}\text{vec}(\Lambda^{'}\Theta) = n\ \text{vec}(S_{xy})^{'}\big(S_y^{-1} \otimes S_x^{-1}\big)\text{vec}(S_{xy}) \\ &= n\ \text{tr}\big(S_y^{-1}S_{yx}S_x^{-1}S_{xy}\big).\end{aligned}$$

In MANOVA, a measure of the strength of association between Θ (multiple traits) and Λ (genotype scores) for the multivariate model $\Theta = \Lambda\text{B} + \text{E}$ depends on a partition of matrix of total SSCP i.e. $\Theta^{'}\Theta = \widehat{\Theta}^{'}\widehat{\Theta} + \text{R}^{'}\text{R}$ [112]. Considering the Pillai-Bartlett (PB) trace, the MANOVA test statistic is stated as $\text{tr}\big(\widehat{\Theta}^{'}\widehat{\Theta}(\Theta^{'}\Theta)^{-1}\big) = \text{tr}\big(\Theta^{'}\Lambda(\Lambda^{'}\Lambda)^{-1}\Lambda^{'}\Theta(\Theta^{'}\Theta)^{-1}\big)$, which can be written in an alternate form $\text{tr}\big(S_y^{-1}S_{yx}S_x^{-1}S_{xy}\big)$. Hence, the GEE-Score test and MANOVA using the PB trace are equivalent.

Muller and Peterson [113] discussed the close relationships among four versions of MANOVA (i.e. with the Pillai-Bartlett trace, Hotelling-Lawley's trace, Wilk's lambda, Roy's largest root), each of which can be written as a function of generalized canonical correlations (CCA). Hence the GEE-Score test is directly related to MANOVA and CCA.

4.2.2.4 Pathway analysis

We extend the adaptive test for association analysis of single-trait and a pathway (i.e. a set of genes) [114] to that of multiple traits and a pathway. The main idea is to allow adaptive weighting at the gene-level, in addition to at the SNP- and trait-levels. Given a pathway G with $|G|$ genes and a single trait t, we partition the score vector according to the genes in G as $U = (U_{1t}', ..., U_{|G|,t}')'$ with a subvector for gene g (with h_g SNPs)

as $U_{gt} = (U_{g,1,t}, ..., U_{g,h_g,t})'$. Denote SPU$(\gamma_1;\ g,\ t)$ and SPUpath$(\gamma_1, \gamma_2;\ t)$ as the gene-specific SPU and the pathway-based SPU test statistics for single trait t, respectively. Define a new test statistic GEE-SPUpath$(\gamma_1, \gamma_2, \gamma_3)$ as the pathway analysis for multiple traits:

$$\text{SPU}(\gamma_1, w_1;\ g,\ t) = \left(\sum_{j=1}^{h_g} (w_{1,g,j} U_{g,j,t})^{\gamma_1} / h_g \right)^{1/\gamma_1},$$

$$\text{SPUpath}(\gamma_1, \gamma_2, w_1, w_2;\ t) = \left(\sum_{g=1}^{|G|} (w_{2,g} \text{SPU}(\gamma_1, w_{1,g};\ g,\ t))^{\gamma_2} \right)^{1/\gamma_2},$$

$$\text{GEE-SPUpath}(\gamma_1, \gamma_2, \gamma_3, w_1, w_2) = \sum_{t=1}^{k} \left(\text{SPUpath}(\gamma_1, \gamma_2, w_1, w_2;\ t) \right)^{\gamma_3},$$

where the three scalars $\gamma_1, \gamma_2, \gamma_3 > 0$ are specified to control the degrees of weighting the SNPs, genes and traits respectively, $w_1 = (w'_{1,1}, ..., w'_{1,|G|})'$ gives gene-specific weights for the SNPs in gene g as $w_{1,g} = (w_{1,g,1}, ..., w_{1,g,h_g})'$, and $w_2 = (w_{2,1}, ..., w_{2,|G|})'$ gives gene-specific weights for each gene in the pathway S. These weights are specified based on some prior knowledge on the importance of the genes and SNPs; without prior knowledge, we can simply use an equal weight 1 on each gene and each SNP, as used in our later simulations. We employed $\gamma_1 \in \Gamma_1 = \{1, 2, ..., 8\}$ and $\gamma_2, \gamma_3 \in \Gamma_2 = \Gamma_3 = \{1, 2, 4, 8\}$ in later simulations.

Finally, a new adaptive test for pathway analysis, denoted GEE-aSPUpath test, is defined as

$$\text{GEE-aSPUpath} = \min_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3} p_{\gamma_1, \gamma_2, \gamma_3},$$

where $p_{\gamma_1, \gamma_2, \gamma_3}$ is the p-value of the GEE-SPUpath$(\gamma_1, \gamma_2, \gamma_3)$ test. The simulation or permutation procedure for generating the null statistics and calculating p-values for all the GEE-SPUpath and GEE-aSPUpath tests are similar to that for the GEE-aSPUset test. We will not discuss the pathway-based tests in the sequel; some simulation results are presented in the Appendix A.

4.3 Results

4.3.1 Real data example

4.3.1.1 GWAS with ADNI-1 data

One objective of ADNI is to elucidate genetic susceptibility to AD. We conducted a gene-based multi-trait analysis for ADNI-1 data, by using grey matter volumes in the 12 ROIs corresponding to the default mode network (DMN) as intermediate phenotypes. DMN is a network of brain regions that are active when an individual is at wakeful rest, which includes inferior temporal, medial orbitofrontal, parahippocampal, precuneus and posterior cingulate ROIs [72]. Importantly, DMN activity distinguishes cognitively impaired patients such as those with Alzheimer's, ADHD, or bipolar disorder from healthy controls [98, 115, 116, 72]. The grey matter volumetric measures related to the DMN were extracted from the ADNI-1 baseline data.

We included all SNPs with minor allele frequency (MAF) ≥ 0.05, genotyping rate more than 90%, and surviving the Hardy-Weinberg equilibrium test at a significance threshold 0.001. After all rounds of quality control, 519,286 SNPs remained, among which 277,527 SNPs were mapped to 17,557 genes. To consider SNPs in promoter or regulatory regions for each gene, we included SNPs upstream and downstream within 20Kb of each gene. Subjects with more than 10% missing genotypes were excluded, and only non-Hispanic Caucasians whose twelve grey matter volumes in DMN were all measured at baseline were included, resulting in 144 AD patients, 311 MCI subjects, and 180 healthy elderly controls. For covariates, gender, years of education, handedness, age, and intracranial volume (ICV) measured at baseline were included.

To demonstrate the applicability and power of our approach, we applied MANOVA, MDMR [107], KMR [26], MFLM [102] and GEE-based tests, GEE-Score, aSPUset and aSPUset-Score tests. The number of MC simulations or permutations for each method was set $B = 10^3$ at beginning, but was increased up to $B = 10^8$ if an obtained p-value was less than $5/B$, which ensured the identification of the genes at the genome-wide significance level (p-value $< 2.8 \times 10^{-6}$ with a Bonferroni adjustment). When any obtained p-value was less than 1.0e-8, we reported it as 1.0e-8. The p-values of permutation-based aSPUset and of simulation-based aSPUset agreed well (with a Pearson correlation 0.98),

thus we reported only permutation-based results. For MFLM, we used beta-smooth basis functions with the Pillai-Bartlett trace as a representative.

The aSPUset and MDMR tests uncovered two loci associated with DMN. Table 4.1 lists the genes with the highest significance levels. *Genes AMOTL1 (on chromosome 11) and APOC1, APOE (on chromosome 19) were identified by both aSPUset and MDMR, but not by other tests, while TOMM40 (on chromosome 19) was only detected by aSPUset.* AMOTL1 is known to be involved in cell adhesion and cell signaling [117]. A recent study using a pathway-enrichment strategy showed that the genes involved in neuronal cell adhesion, and cell signaling are overrepresented in schizophrenia and bipolar disorder [115]. Anney et al. [118] identified AMOTL1 as a gene associated with ADHD. The gene was also highly expressed in thalamus, a brain region implicated in the cognitive impairment of early stage Huntington's disease [119]. Three genes (APOC1, APOE, TOMM40) in chromosome 19 could not be readily discerned due to their physical closeness, though their gene sizes (i.e. the numbers of SNPs) varied. The p-values of MDMR became less significant as the gene size increased, while the aSPUset was robust to the number of SNPs. This locus containing APOE is well known to be related to Alzheimer's disease and cognitive impairment disorder [120, 121, 122].

Table 4.2 lists the SNPs included in the significant genes. We applied several single SNP-based tests for association with the default mode network. For each method, the permutation or simulation number was increased up to 10^8 to satisfy the genome-wise significance level. As shown in Table 4.2, none of the SNPs in gene AMOTL1 was significant, suggesting that a strong association signal was retained only in the gene-level, rather than in the SNP-level. On the other hand, SNP rs429358 contained in three genes (APOC1, APOE, TOMM40) was highly significant with p-value of 1.0e-8. *These results lend support for the proposed aSPUset test's potential of being able to recover both multiple weak effects and single strong effects, due to its adaptiveness.*

We explored each identified locus in details in Figures 4.1 and 4.2. In Figure 4.1, a LocusZoom plot [87] illustrates local linkage disequilibrium (LD), recombination patterns and p-values obtained from the single SNP-based aSPU test for DMN. Figure 4.2 illustrates the association analyses for genes AMOTL1 and APOE respectively. First we obtained p-values from the univariate test between each SNP and each individual trait comprising DMN, then applied SNP-based test (aSPU) between each SNP and DMN

(12 traits). Finally, we applied the aSPUset test at the gene level for DMN. The SNPs contained in AMOTL1 showed strong LD (Figure 4.1A), and their aggregate effects turned out to be significant at the gene level (Figure 4.2A). Among the SPU(γ_1, γ_2) tests applied with $\gamma_1, \gamma_2 \in \{1, ..., 8, \infty\}$, SPU(3,2) showed the minimum p-value, implying that weak effects were aggregated for an overall association. In Figure 4.2B, only one variant (rs429358) in APOE was significant, but the significance level of aSPUset did not diminish in the gene level analysis. In testing APOE, the p-values of SPU(2,1), SPU(4,1), SPU(6,1), SPU(8,1), and SPU(∞,1) were tied and the most significant; this suggested that one SNP (rs429358) dominated across in the gene level across all the traits.

Since the proposed test is based on combining all possible single SNP-single trait association pairs, if one would like to identify which pairs contribute most to an overall association, one can simply examine the signifcance levels of the univariate single SNP-single trait association tests. For example, Figure 4.2 (left panels) illustrates the contribution of each SNP-trait pair for AMOTL1 and APOE. In the gene AMOTL1, the SNP-trait pairs, (rs1367505, R-InferiorTemporal), (rs2033367, R-InferiorTemporal) and (rs333027, L-InferiorParietal), were ranked highest; for APOE, the top 3 significant pairs were (rs429358 , R-Precuneus), (rs2075650, L-Precuneus) and (rs429358, L-InferiorParietal).

We conducted a single SNP-based GWAS scan for the ADNI-1 data. *Interestingly, no SNP was significant from univariate single SNP–single trait analyses (not shown). Furthermore, only one SNP, rs429358, was significant in single SNP-based multi-trait analyses. In contrast, two loci (AMOTL1 and APOE) were uncovered by gene-based multi-trait analyses by our proposed new test and MDMR.* In all analyses, covariates considered included gender, years of education, handedness, age, and intracranial volume (ICV) measured at baseline. *Taken together, these results clearly demonstrated the advantage and power gain of our proposed gene-based multi-trait analysis.*

4.3.1.2 Validation with ADNI-GO/2 data

Using the ADNI-1 data as the discovery sample, our GWAS identified two loci associated with DMN. To validate the results, each method was applied to the two genes AMOTL1 and APOE using the ADNI-GO/2 data as the validation sample (with $n = 754$). We

applied the same SNP-filtering criteria as applied to ADNI-1. Table 4.3 presents the p-values obtained from each method; no significant association was identified. Due to different genotyping arrays, ADNI-GO/2 data contains different sets of SNPs from those of ADNI-1; we imputed missing SNPs which were originally included in the analysis of ADNI-1, based on the reference samples of HapMap 3 with MaCH [123], in order to apply each method to the identical SNP sets of ADNI-1. The aSPUset and aSPUset-Score tests identified gene APOE with p-values 0.019 and 0.024 respectively, which passed the significance threshold 0.05/2 as shown in Table 4.3, but gene AMOTL1 was not significant by any test.

We would mention possible sample differences between ADNI-1 and ADNI-GO/2 cohorts. The ADNI-1 cohort includes three subject groups consisting of 25% AD patients, 50% subjects with MCI (Mild Cognitive Impairment) and 25% CN (Cognitively Normal) subjects; in contrast, the ADN-GO/2 study assigns 754 subjects into five groups: 20% CN , 12% SMC (Significant Memory Concern), 35% EMCI (Early Mild Cognitive Impairment), 17% LMCI (Late Mild Cognitive Impairment), and 16% AD. At least the proportions of the CN subjects and AD patients in the two cohorts are different, which might lead to different association results.

Finally, we combined the two cohorts to form ADNI-1/GO/2 with a larger sample size (about 1400 subjects) and obtained the p-values from the tests for the two candidate gene regions. The two genes were highly significantly associated with the default mode network as shown in Table 4.3.

4.3.1.3 Gene-based rare variant analysis of the ADNI sequencing data

The proposed method was applied to analysis of rare variants with the ADNI whole-genome sequencing (WGS) data, consisting of 254 and 500 subjects from ADNI-1 and ADNI-GO/2 respectively. In total, 26,142 genes were included for analyses; all variants inside a gene and those located 25kb of upstream and downstream of the gene were mapped to the gene. Five covariates were adjusted: gender, years of education, handedness, age and ICV. Due to the low frequency of rare variants, the asymptotic assumption for some tests may not hold; we modified each method to avoid using asymptotics. For MANOVA, rather using the usual F-distribution, we permuted residuals (under the null model) to estimate its null distribution; for aSPUset and MFLM, similarly the

permutation-based method was applied. We included all rare variants within each gene region; the number of variants within each region ranged from 3 to 750. Sometimes permutation-based MANOVA suffered from rank deficiency when constructing the test statistic and could not be applied to about 600 genes; MFLM also failed for some genes due to rank deficiency.

First we included only rare variants (with MAF < 0.01), then both rare and low-frequency variants (with MAF < 0.05). No gene passed the genome-wide Bonferroni-adjusted significance threshold of 2.8×10^{-6}. MFLM was problematic with an inflation factor around 1.5 in both analyses.

Given that two gene regions were significantly associated with DMN in the previous GWAS analysis, it would be of interest to see whether the rare variants in the two genes were associated. Table 4.4 reports the p-values for the two candidate genes. No significant associations were detected.

4.3.2 Simulations

4.3.2.1 Simulation set-ups

We evaluated the performance of our method along with several existing methods in simulation studies. The simulated data mimicked the association structures for the two genes (AMOTL1 on chromosome 11 and TOMM40 on chromosome 19) and default mode network (DMN) in ADNI-1 data. Two factors were considered: association effect size (Set-up 1) and sparsity of association patterns (Set-up 2). For Set-up 1, various effect sizes were created by scaling the regression coefficient estimates obtained from a multivariate linear model (MLM) fitted to the original data. On each gene, an MLM was fitted to the ADNI-1 data, including the covariates (z_i), SNPs (x_i) and DMN (Y_i). For covariates, we included gender, education, handedness, age, and ICV as in the original data analysis. Denote the parameter estimates in an MLM as follows: G_0 is a vector for intercepts; $G = (g_{jt})$ is a $p \times k$ matrix, in which g_{jt} represents the effect size of SNP j on trait t; the element h_{qt} in matrix $H = (h_{qt})$ stands for the qth covariate effect on the tth trait; Σ is the covariance estimate for the multivariate error term. To maintain the true correlation structures among genotype scores $x_i = (x_{i1}, ..., x_{ip})'$ and five covariates $z_i = (z_{i1}, ..., z_{i5})'$, we sampled pairs (x_i, z_i) from the ADNI-1 data in each simulation.

The multiple traits for subject i were generated from a multivariate normal distribution:

$$Y_i \sim \mathcal{MN}(G_0 + \phi \cdot G' x_i + H' z_i, \ \ \Sigma). \tag{4.12}$$

Here ϕ was a scaling parameter controlling the effect sizes of the SNPs (x_i): with $\phi = 0$, the null hypothesis held and Type I error rates were evaluated; at $\phi = 1$, the effect sizes were set to be equal to the estimated ones from the ADNI-1 data.

For Set-up 2, we varied the sparsity level of the association structure. At a fixed $\phi = 0.5$, we increased the gene size by adding some null SNPs to gene AMOTL1. For the null SNPs, the genotype data adjacent to AMOTL1 were used. As before, (x_i, z_i) pairs were sampled from the ADNI-1 data. Throughout simulations, 10000 replicates were used for each set-up and the tests were conducted at the significance level $\alpha = 0.05$.

4.3.2.2 Type I error and power

All the tests showed Type I error rates controlled under the nominal level 0.05 (Table 4.5). Of note, MDMR resulted in conservative Type I error rates. In Set-up 1 (Table 4.5), as the association effect size (ϕ) decreased, the aSPUset and aSPUset-Score tests were more powerful than other tests, suggesting the potential usefulness of the proposed tests in identifying causal SNPs with weak effects. Since MFLM was proposed to reduce the dimensionality of the SNP data, it might not be desirable to use MFLM here; it might perform better with larger numbers of SNPs.

In Set-up 2 (Table 4.6), the aSPUset and aSPUset-Score yielded higher power than other tests as the proportion of the null SNPs in the SNP set increased. Throughout the simulations, the GEE-Score test performed similarly to MANOVA, confirming their equivalence.

4.3.2.3 Computational time

We reported computational requirement of each method in Table 4.7 by taking the average computation time for simulation Set-up 2. MANOVA was computationally most efficient, followed by MFLM. As the number of SNPs increased, GEE-Score test and aSPUset-Score test became computationally more demanding, but still feasible.

4.4 Conclusions

We have presented a highly adaptive association test for multiple traits and multiple genetic variants. From the GWAS analyses of the ADNI-1 data, we observed its potential power gains in identifying cumulative weak effects of multiple associated SNPs in gene AMOTL1 with multiple traits, which were undetectable by several other gene-based tests and single SNP-based tests. Given that most common variants have only weak effects for complex diseases and traits, developing testing strategies to improve power in identifying multiple SNPs with weak effects is very important. Our proposed method is developed along this direction. Furthermore, due to its adaptiveness, it also retains power in the presence of only one or few associated SNPs (or traits), as shown for the APOE gene with the ADNI-1 data (while several existing gene-based tests failed to capture). Our proposed adaptive test is in contrast to most of the existing tests, which may be powerful in one or more situations, but not across a wide range of situations. In practice, since the true association pattern for a given gene and traits is unknown, it is unclear which non-adaptive test should be used; it will be convenient and promising to apply an adaptive test such as our proposed one.

We emphasize the potential power gain with the use of multiple traits, especially of intermediate phenotypes for a complex disease such as AD [105, 124]. However, since it is unknown how many of, and in what association patterns, the multiple traits are associated with a gene (or a set of SNPs), a straightforward use of any multivariate test may lose, not gain, power. Again, the availability of a powerful and adaptive test such as our proposed one will largely facilitate its easy and effective use in practice.

Finally, we summarize the use of our proposed tests and make some recommendations. To assess an overall association between a set of SNPs and a set of traits, we would recommend the use of the p-value of the aSPUset test. If it is significant, one can check the individual p-values of the SPU(γ_1, γ_2) tests to shed some light on the underlying association pattern. If a larger γ_1 (or γ_2) leads to a more signifcant p-value of the SPU test, it would suggest a more sparse association pattern; that is, perhaps one a fewer number of the SNPs (or traits) are associated. Furthermore, one can examine the p-value from the univariate test for each SNP-trait pair to identify which SNP-trait pairs contribute most to the overall association. For choosing candidate values of γ_1

and γ_2, based on our limited experience, we suggest using $\Gamma_1 = \Gamma_2 = \{1, 2, ..., 8, \infty\}$ by default, though an optimal choice depends on the situation; using a too large or too small set Γ_1 or Γ_2 will lead to loss of power. A general guidance, taking Γ_1 as an example (and similarly for Γ_2), is to use $\Gamma_1 = \{1, 2, ..., C_1, \infty\}$ such that the SPU(C_1, γ_2) test gives a p-value almost equal to that of SPU(∞, γ_2); a larger number of SNPs may require a larger value of C_1. In addition, if some large univariate associations between various SNP-trait pairs are likely to be in opposite directions, only even integers are needed in Γ_1 and Γ_2; if it is known a priori that large univariate associations are mainly in one direction, then using only odd integers may be most powerful; otherwise, both even and odd integers should be used. Given the relationships among the tests, we recommend the use of our proposed aSPUset and aSPUset-Score tests, though MFLM may also perform well for large genes; further evaluations are needed.

Table 4.1: P-values of the gene-based association tests for DMN with the ADNI-1 data.

					GEE						
Gene-region	#SNPs	Chr	Position		Score	aSPUset	aSPUset-Score	MANOVA	MDMR	KMR	MFLM
AMOTL1	6	11	94121155	94269566	1.18e0-4	**1.0e-08**	**1.0e-08**	7.73e-05	**3.48e-07**	0.451	7.73e-05
APOC1	4	19	50089760	50134446	6.14e-04	**1.0e-08**	**1.0e-08**	3.45e-04	**4.42e-08**	0.342	2.30e-04
APOE	6	19	50080878	50124490	1.27e-03	**1.0e-08**	**1.0e-08**	7.93e-04	**2.21e-07**	0.268	5.97e-04
TOMM40	10	19	50066316	50118786	0.023	**1.0e-08**	**1.0e-08**	1.86e-02	6.99e-06	0.569	1.04e-03

Table 4.2: P-values of the single SNP-based association tests for DMN for the significant gene-regions (±20kb) with the ADNI-1 data.

					GEE					
Gene	Chr	aSPUset	SNP	Position	Score	SPU(2)	SPU(∞)	aSPU	MANOVA	MDMR
AMOTL1	11	1.0e-08	rs1367505	94186285	8.0e-05	2.4e-07	2.8e-05	5.1e-07	5.1e-05	2.1e-07
			rs10501816	94187396	0.417	0.151	0.237	0.158	0.432	0.186
			rs2033367	94195356	1.2e-04	8.0e-07	6.5e-05	1.6e-06	9.1e-05	3.01e-07
			rs2241667	94203379	8.0e-04	1.6e-06	1.3e-04	3.9e-06	1.8e-04	8.0e-06
			rs333027	94225561	5.0e-04	1.6e-05	9.5e-05	3.1e-05	4.6e-04	6.9e-05
			rs333025	94227040	0.02	0.025	0.030	0.045	0.015	0.022
APOC1	19	1.0e-08	rs8106922	50093506	0.236	0.116	0.212	0.183	0.244	0.128
			rs405509	50100676	0.420	0.156	0.207	0.186	0.422	0.184
			rs439401	50106291	7.0e-04	2.3e-06	1.2e-05	3.1e-06	4.1e-04	2.2e-05
			rs429358	50103781	1.0e-05	**4e-0**8	8.3e-06	**1.0e-08**	2.1e-06	**1.25e-08**
APOE	19	1.0e-08	rs157580	50087106	3.1e-03	1.4e-04	8.8e-04	9.0e-05	3.1e-03	3.9e-4
			rs2075650	50087459	9.0e-04	3.8e-06	2.2e-03	1.2e-06	2.9e-04	1.5e-05
			rs8106922	50093506	0.236	0.116	0.212	0.183	0.244	0.128
			rs405509	50100676	0.420	0.156	0.207	0.186	0.422	0.184
			rs439401	50106291	7.0e-04	2.3e-06	1.2e-05	3.1e-06	4.1e-04	2.2e-05
			rs429358	50103781	1.0e-05	**4e-0**8	8.3e-06	**1.0e-08**	2.1e-06	**1.25e-08**
TOMM40	19	1.0e-08	rs2075642	50069307	0.842	0.711	0.471	0.629	0.840	0.662
			rs387976	50070900	0.073	0.031	0.036	0.040	0.068	0.067
			rs11667640	50071631	0.262	0.034	0.012	0.021	0.265	0.035
			rs6859	50073874	0.728	0.076	0.299	0.057	0.729	0.072
			rs157580	50087106	3.1e-03	1.4e-04	8.8e-04	9.0e-05	3.1e-03	3.9e-4
			rs2075650	50087459	9.0e-04	3.8e-06	2.2e-03	1.2e-06	2.9e-04	1.5e-05
			rs8106922	50093506	0.236	0.116	0.212	0.183	0.244	0.128
			rs405509	50100676	0.420	0.156	0.207	0.186	0.422	0.184
			rs439401	50106291	7.0e-04	2.3e-06	1.2e-05	3.1e-06	4.1e-04	2.2e-05
			rs429358	50103781	1.0e-05	**4e-0**8	8.3e-06	**1.0e-08**	2.1e-06	**1.25e-08**

Table 4.3: P-values of the gene-based association tests with the ADNI-GO/2 and ADNI-1/GO/2 data.

						GEE					
Data	Gene-region	#SNPs	Chr	Position		Score	aSPUset	aSPUset-Score	MANOVA	MDMR	MFLM
ADNI-GO/2	AMOTL1	13	11	94481507	94629918	0.723	0.896	0.940	0.698	0.716	0.638
	APOE	13	19	45389277	45432652	0.083	0.042	0.056	0.097	0.366	0.974
ADNI-GO/2 with	AMOTL1	6	11	-		0.639	0.552	0.576	0.638	0.918	0.638
identical SNP sets of ADNI-1	APOE	6	19	-		0.308	**0.019**	**0.024**	0.292	0.065	0.292
ADNI-1/GO/2 with	AMOTL1	6	11	-		**1.0e-08**	**1.0e-08**	**1.0e-08**	**1.0e-08**	**1.0e-08**	**1.0e-08**
identical SNP sets of ADNI-1	APOE	6	19	-		**1.0e-08**	**1.0e-08**	**4.45e-06**	**1.0e-08**	**1.0e-08**	**4.45e-06**

Table 4.4: P-values of the gene-based tests for rare variant–DMN association with the ADNI sequencing data.

Filtering criteria	Gene-region	# SNPs	Chr	Position		aSPUset	MANOVA	MFLM
MAF< 0.05	AMOTL1	536	11	94481507	94629918	0.298	0.176	0.148
	APOE	153	19	45389277	45432652	0.104	0.837	0.476
MAF< 0.01	AMOTL1	265	11	94481507	94629918	0.835	0.193	0.151
	APOE	84	19	45389277	45432652	0.874	0.833	0.189

Table 4.5: Simulation setup 1: Type I errors ($\phi = 0$) and power ($\phi \neq 0$) under varying genetic effect sizes.

AMOTL1 (6 SNPs)

	GEE						
ϕ	Score	SPU(2,2)	aSPUset	aSPUset-Score	MANOVA	MDMR	MFLM
0	0.0479	0.0528	0.0530	0.0522	0.0490	0.0353	0.0490
0.2	0.1078	0.1837	0.1659	0.1654	0.1128	0.0964	0.1128
0.3	0.2325	0.3494	0.3159	0.3328	0.2394	0.2135	0.2394
0.4	0.4657	0.5571	0.5079	0.5559	0.4764	0.4130	0.4764
0.5	0.7436	0.7614	0.7156	0.7967	0.7528	0.6607	0.7528
0.6	0.9288	0.9008	0.8722	0.9452	0.9341	0.8608	0.9341
0.7	0.9913	0.9677	0.9550	0.9926	0.9921	0.9611	0.9921

TOMM40 (10 SNPs)

	GEE						
ϕ	Score	SPU(2,2)	aSPUset	aSPUset-Score	MANOVA	MDMR	MFLM
0	0.0488	0.0483	0.0482	0.0495	0.0505	0.0323	0.0532
0.2	0.1051	0.1719	0.1347	0.1369	0.1110	0.0903	0.1116
0.3	0.2177	0.3643	0.2763	0.2889	0.2262	0.2053	0.2169
0.4	0.4429	0.6121	0.5018	0.5330	0.4605	0.4246	0.4256
0.5	0.5800	0.7304	0.6231	0.6673	0.5958	0.5593	0.5664
0.6	0.7196	0.8271	0.7369	0.7904	0.7346	0.6885	0.7036
0.7	0.8405	0.8983	0.8293	0.8856	0.8489	0.8015	0.8231

Table 4.6: Simulation setup 2: power under varying sparsity levels of association pattern.

AMOTL1+ Null SNPs

# total	# causal	# null	GEE					
SNPs	SNPs	SNPs	Score	aSPUset	aSPUset-Score	MANOVA	MDMR	MFLM
6	6	0	0.7436	0.7156	0.7967	0.7528	0.6607	0.7528
12	6	6	0.5332	0.6495	0.6923	0.5427	0.4904	0.5228
18	6	12	0.4160	0.6149	0.6336	0.4291	0.3884	0.3882
30	6	24	0.2950	0.4495	0.4617	0.3055	0.2819	0.2872
60	6	54	0.1813	0.3120	0.3150	0.1981	0.1756	0.2124
80	6	74	0.1442	0.2912	0.2912	0.1661	0.1434	0.1697

Table 4.7: Mean computing times (in seconds) for simulation setup 2.

# total	GEE					
SNPs	Score	aSPUset	aSPUset-Score	MANOVA	MDMR	MFLM
12	1.1597	1.2472	1.6261	0.0149	24.2924	0.0354
18	1.3398	1.5062	2.2552	0.0156	22.2903	0.0385
30	2.2541	1.8766	3.7482	0.0172	21.5940	0.0449
60	6.5183	2.8785	11.1315	0.0211	19.3995	0.0612
80	11.8868	3.5546	20.4237	0.0243	18.4600	0.0722

Figure 4.1: LocusZoom for two loci identified by aSPUset and MDMR: LD structure in each locus and p-values obtained from the single SNP-based aSPU test are presented.

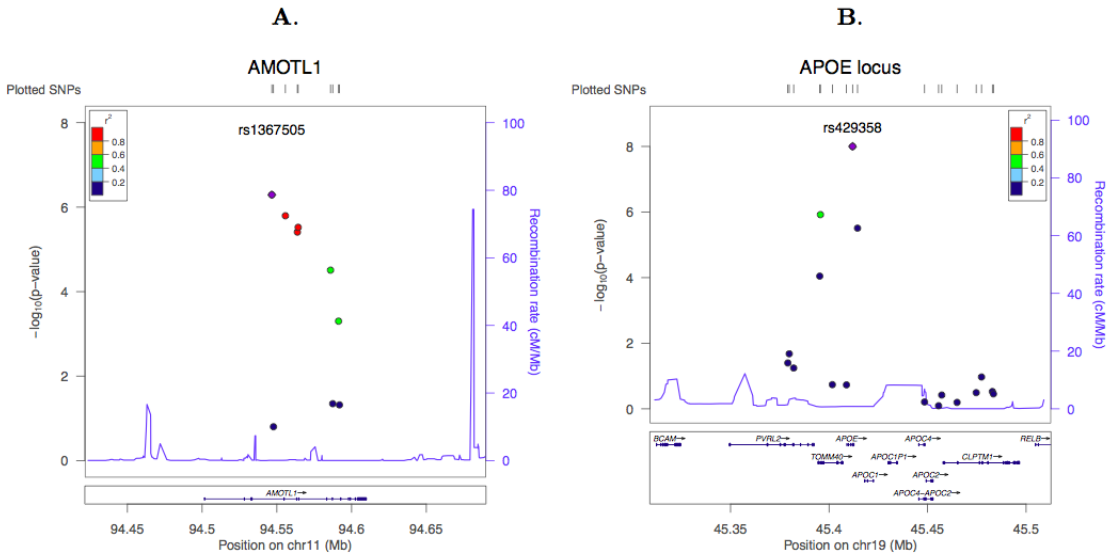
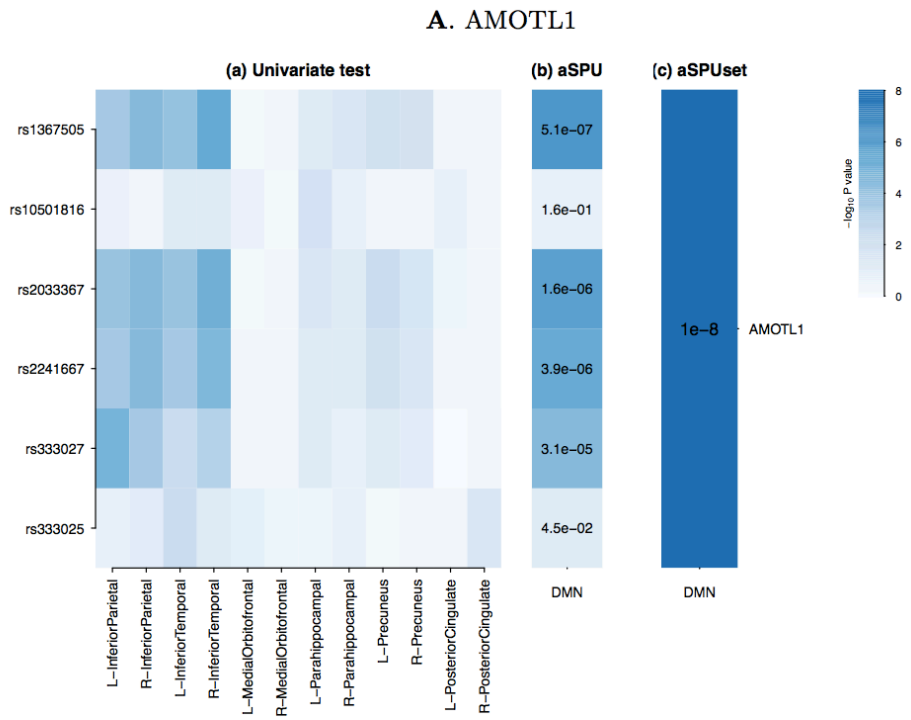
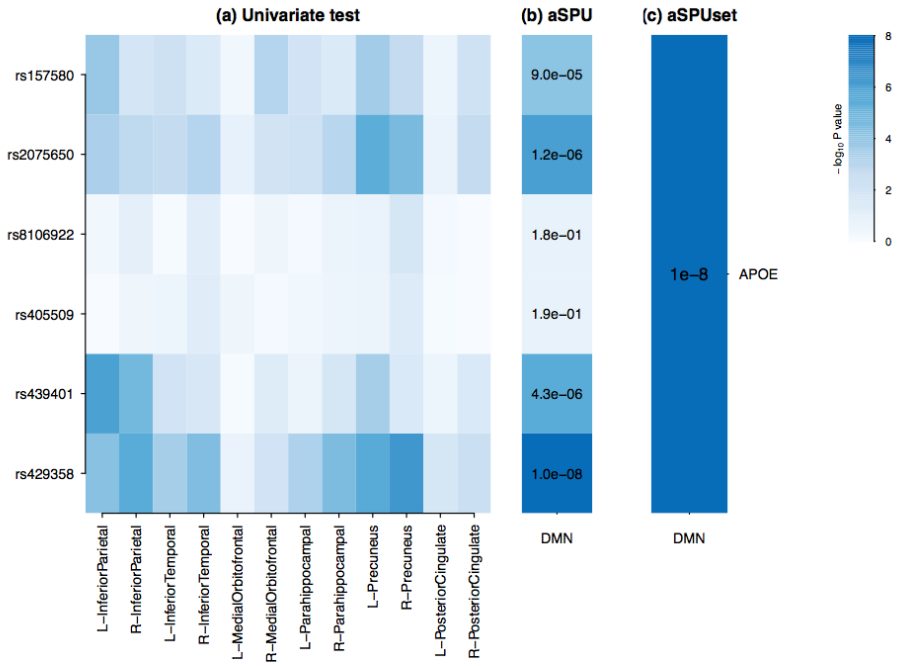


Figure 4.2: P-values of the association tests for DMN and SNPs for genes *AMOTL1* and *APOE*: (a) univariate test for single SNP–single trait association; (b) aSPU test for single SNP–multitrait association; (c) aSPUset for gene–multitrait association.





Chapter 5

Discussion and future work

Many studies have paved the way for genetic association analyses using secondary imaging phenotypes [8, 9, 10, 11]. Though in general using a standard linear regression for secondary phenotypes will lead to biased inference, the current practice in neuroimaging genetics does not consider this potential problem. In Chapter 2, we addressed whether using standard linear regression of secondary phenotypes would lead to biased estimates. Under practical situations mimicking the ADNI data, the standard linear regression method without any adjustment (unadj-lm), but not the one adjusting for the disease status (D-adj-lm), performed similarly to two valid methods, an inverse probability weighted regression [13] and a method using a retrospective likelihood [14]. We assumed that it was because the AD prevalence in the age-matched population was similar to the sampling proportion of the cases in ADNI data, leading to small biases. The illustrative example (Section 2.3.3) demonstrated that unadj-lm and D-adj-lm were likely to yield different results, because one targets a marginal association, while the other models a conditional association. Two valid methods [13, 14] drew correct inferences in all applications, but they were either inefficient or not easy to use. Finally we concluded that, under the situations similar to the ADNI data (where the disease prevalence in cohort samples resembles that of the target population), using standard linear regression without an adjustment would not cause any severe problem, though one may still need cautions. There are limited statistical methods to handle SNP-secondary phenotype associations for complex situations beyond the case-control design; it is challenging but could be interesting topics to be studied.

Motivated by Chapter 2, we proposed two novel association tests for multiple imaging phenotypes. Both tests (Chapter 3 and Chapter 4) were built on the adaptive sum of powered score (SPU) test [28]: Chapter 3 presented a new adaptive test for single SNP-multi trait associations in a POM. Different from Zhang et al. [28], the proposed method accommodates a non-linear relationship between the genotype and the phenotypes, is robust to the assumed inheritance mode, and is computationally more efficient than Zhang et al. [28], but often showing similar performance to that of Zhang et al. The method is applicable to high dimensional phenotypes as we applied to brain functional connectivity as phenotypes (Section 3.3.1.2). Compared to the existing dimension reduction methods for high dimensional phenotypes [21, 22, 23, 24, 61, 62], the proposed POM based adaptive test was able to detect associations caused by joint weak effects of individual traits as demonstrated in Section 3.3.1.2. The method can be extended to gene- or pathway- based analysis, which is definitely of interest in the future works.

Chapter 4 introduced a statistical method for multi-SNP and multi-trait associations by extending GEE-based adaptive test [28]. We called the novel tests as GEE-aSPUset and GEE-pathway test for gene- and pathway- based test respectively. The GEE-aSPUset test is adaptive to varying association patterns at a SNP level and at a trait level, while the GEE-pathway test is adaptive to possibly varying gene-level associations in addition to the SNP level and the trait level. We compared the performance of the new test (GEE-aSPUset) with several existing tests using both simulated and real data. Genes $AMOTL1$ on chromosome 11 and $APOE$ on chromosome 19 were discovered by the new test to be significantly associated with default mode network. Notably, gene $AMOTL1$was not detected by single SNP-based analyses. One SNP rs429358 contained in $APOE$ was highly significant. These results lend support for the proposed aSPUset test's potential of being able to recover both multiple weak effects and single strong effect, due to its adaptiveness. We also illustrated the application of GEE-aSPUset test to the ADNI whole-genome sequencing (WGS) data, though none significant genes were identified, presumably due to a relatively small sample size. For the future work, we can consider to incorporate the high dimensional phenotype in the GEE framework. Current proposed works (GEE-aSPUset and GEE-aSPUpath tests) are only applicable to a large sample setting where the sample size should be larger than the number of phenotypes; a penalized GEE procedure [125, 126] would be a useful strategy for the

application of the proposed tests.

References

[1] K.L. Bigos, A.R. Hariri, and D.R. Weinberger. *Neuroimaging Genetics: Principles and Practices*. Oxford University Press, 2016.

[2] L. Shen, P.M. Thompson, S.G. Potkin, L. Bertram, L.A. Farrer, T.M. Foroud, R.C. Green, X. Hu, et al. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav.*, 8 (2):183–207, 2014.

[3] C.M. Karch, C. Cruchaga, and A.M. Goate. Alzheimer's disease genetics: from the bench to the clinic. *Neuron*, 83 (1):11–26, 2014.

[4] P.G. Ridge, S. Mukherjee, P.K. Crane, J.S. Kauwe, et al. Alzheimer's disease: analyzing the missing heritability. *PLoS ONE*, 8:e79771, 2013.

[5] O.L. Lopez, W.J. Jagust, S.T. DeKosky, J.T. Becker, A. Fitzpatrick, C. Dulberg, et al. Prevalence and classification of mild cognitive impairment in the cardiovascular health study cognition study. *Arch. Neurol.*, 60:1385–1389, 2003.

[6] R.O. Roberts, Y.E. Geda, D.S. Knopman, R.H. Cha, V.S. Pankratz, B.F. Boeve, et al. The Mayo clinic study of aging: design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*, 30:58–69, 2008.

[7] T. Hanninen, M. Hallikainen, S. Tuomainen, M. Vanhanen, and H. Soininen. Prevalence of mild cognitive impairment: a population-based study in elderly subjects. *Acta Neurol. Scand.*, 106:148–154, 2002.

[8] L. Shen, S. Kim, S.L. Risachera, K. Nho, S. Swaminathan, J.D. Westa, T. Foroudd, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. *NeuroImage*, 53 (3):1051–1063, 2010.

[9] J.L. Stein, X. Hua, S. Lee, A.J. Ho, A.D. Leow, A.W. Toga, A.J. Saykin, L. Shen, T. Foroud, et al. Voxelwise genome-wide association study (vGWAS). *NeuroImage*, 53 (3):1160–1174, 2010a.

[10] S.A. Meda, B. Narayanan, J. Liu, N.I. Perrone-Bizzozero, M.C. Stevens, V.D. Calhoun, et al. A large scale multivariate parallel ICA method reveals novel imaging-genetic relationships for Alzheimer's disease in the adni cohort. *NeuroImage*, 60 (3):1608–1621, 2012.

[11] D.P. Hibar, J.L. Stein, N. Jahanshad, O. Kohannim, X. Hua, et al. Genome-wide interaction analysis reveals replicated epistatic effects on brain structure. *Neurobiol. Aging*, 36 (Suppl. 1):S151–S158, 2015b.

[12] E.D. Schifano, L. Li, D.C. Christiani, and X. Lin. Genome-wide association analysis for multiple continuous secondary phenotypes. *Am. J. Hum. Genet.*, 92:744–759, 2013.

[13] G.M. Monsees, R.M. Tamimi, and P. Kraft. Genome-wide association scans for secondary traits using case-control samples. *Genet. Epidemiol.*, 33:717–728, 2009.

[14] D.Y. Lin and D. Zeng. Proper analysis of secondary phenotype data in case-control association studies. *Genet. Epidemiol.*, 33:256265, 2009.

[15] S. Van der Sluis, D. Posthuma, and C.V. Dolan. TATES: efficient multivariate genotype- phenotype analysis for genome-wide association studies. *PLoS Genet.*, 9:e1003235, 2013.

[16] S. van der Sluis, CV. Dolan, J. Li, Y. Song, et al. MGAS: a powerful tool for multivariate gene-based genome-wide association analysis. *Bioinformatics*, 31 (7):1007–1015, 2015.

[17] X. Guo, Z. Liu, X. Wang, and H. Zhang. Genetic association test for multiple traits at gene level. *Genet. Epidemiol.*, 37 (1):122–129, 2013.

[18] K. Wang and D. Abbott. A principal components regression approach to multilocus genetic association studies. *Genet. Epidemiol.*, 32:108–118, 2007.

[19] L. Klei, D. Luca, B. Devlin, and K. Roeder. Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genet. Epidemiol.*, 32:9–19, 2008.

[20] M.A. Ferreira and S.M. Purcell. A multivariate test of association. *Bioinformatics*, 25:132133, 2009.

[21] M. Vounou, T.E. Nichols, G. Montana, and Alzheimer's Disease Neuroimaging Initiative. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage*, 53 (3):1147–1159, 2010.

[22] J. Lin, H. Zhu, R. Knickmeyer, M. Styner, J. Gilmore, and J.G. Ibrahim. Projection regression models for multivariate imaging phenotype. *Genet. Epidemiol.*, 36:631–641, 2012.

[23] Q. Sun, H. Zhu, Y. Liu, J.G. Ibrahim, and Alzheimer's Disease Neuroimaging Initiative. Sprem: Sparse projection regression model for high-dimensional linear regression. *J. Am. Stat. Assoc.*, 110 (509):289–302, 2015.

[24] L. Du, H. Huang, J. Yan, S. Kim, S.L. Risacher, et al. Structured sparse canonical correlation analysis for brain imaging genetics: an improved GraphNet method. *Bioinformatics*, 2016.

[25] D.Y. Lin and Z.Z. Tang. A general framework for detecting disease associations with rare variants in sequencing studies. *Am. J. Hum. Genet.*, 89:354–367, 2011.

[26] A. Maity, P.F. Sullivan, and J.Y. Tzeng. Multivariate phenotype association analysis by marker-set kernel machine regression. *Genet. Epidemiol.*, 36:686–695, 2012.

[27] X. Wang, S. Lee, X. Zhu, and X. Redline, S.and Lin. Gee-based snp set association test for continuous and discrete traits in family-based association studies. *Genet. Epidemiol.*, 37:778–786, 2013.

[28] Y. Zhang, Z. Xu, X. Shen, W. Pan, and Alzheimer's Disease Neuroimaging Initiative. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*, 96:309–325, 2014.

[29] T.A. Manolio, F.S. Collins, N.J. Cox, D.B. Goldstein, et al. Finding the missing heritability of complex diseasesn. *Nature*, 461:747–753, 2009.

[30] K. Liang and S. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.

[31] J. Kim and W. Pan. A cautionary note on using secondary phenotypes in neuroimaging genetic studies. *Neuroimage*, 121:136–145, 2015.

[32] J. Kim, Y. Zhang, and W. Pan. Powerful and adaptive testing for multi-trait and multi-snp associations with GWAS and sequencing data. *Genetics*, 2016.

[33] D.J. Hunter, P. Kraft, K.B. Jacobs, D.G. Cox, M. Yeager, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmeno-pausal breast cancer. *Nat. Genet.*, 39:870–874, 2007.

[34] L.J. Scott, K.L. Mohlke, L.L. Bonnycastle, C.J. Willer, Y. Li, W.L. Duren, M.R. Erdos, et al. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*, 316:1341–1345, 2007.

[35] G. Thomas, K.B. Jacobs, M. Yeager, P. Kraft, S. Wacholder, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.*, 40:310–315, 2008.

[36] R.L. Prentice and R. Pyke. Logistic disease incidence models and case-control studies. *Biometrika*, 66 (3):403–411, 1979.

[37] J. Wei, R.J. Carroll, U.U. Muller, and I.V. Keilegom. Robust estimation for homoscedastic regression in the secondary analysis of case-control data. *J. R. Stat. Soc. B*, 75:185–206, 2013.

[38] A. Ghosh, F. Wright, and F. Zou. Unified analysis of secondary traits in case-control association studies. *J. Am. Stat. Assoc.*, 108 (502):566–576, 2013.

[39] H.Y. Chen, R. Kittles, and W. Zhang. Bias correction to secondary trait analysis with case-control design. *Stat. Med.*, 32 (9):1494–1508, 2013.

[40] E.J.T. Tchetgen. A general regression framework for a secondary outcome in case-control studies. *Biostatistics*, 5 (1):117–128, 2014.

[41] S.G. Potkin, F. Macciardi, G. Guffanti, J.H. Fallon, Q. Wang, J.A. Turner, A. Lakatos, et al. Identifying gene regulatory networks in schizophrenia. *NeuroImage*, 53 (3):839–847, 2010.

[42] D.P. Hibar, J.L. Stein, and M.E. Renteria. Common genetic variants influence human subcortical brain structures. *Nature*, 520:224–229, 2015a.

[43] D.B. Richardson, P. Rzehak, J. Klenk, and S.K. Weiland. Analyses of casecontrol data for additional outcomes. *Epidemiology*, 18:441–445, 2007.

[44] S.M. Lutz, J.E. Hokanson, and C. Lange. An alternative hypothesis testing strategy for secondary phenotype data in case-control genetic association studies. *Front. Genet.*, 5:188, 2014.

[45] Alzheimer's Association. Alzheimer's Disease Facts and Figures. *Alzheimer's and Dementia*, 8(2), 2012.

[46] Alzheimer's Association. Alzheimer's Disease Facts and Figures. *Alzheimer's and Dementia*, 10(2), 2014.

[47] L.E. Hebert, P.A. Scherr, J.L. Bienias, D.A. Bennett, and D.A. Evans. Alzheimer disease in the U.S. population: prevalence estimates using the 2000 Census. *Arch. Neurol.*, 60 (8):1119–1122, 2013.

[48] D.H. Kim, M.E. Payne, R.M. Levy, J.R. MacFall, and D.C. Steffens. APOE genotype and hippocampal volume change in geriatric depression. *Biol. Psychiatry*, 51 (5):426–429, 2002.

[49] P.H. Lu, P.M. Thompson, A. Leow, G.J. Lee, A. Lee, I. Yanovsky, et al. Apolipoprotein E genotype is associated with temporal and hippocampal atrophy rates in healthy elderly adults: a tensor-based morphometry study. *J. Alzheimers Dis.*, 23 (3):433–442, 2011.

[50] E. Mori, K. Lee, M. Yasuda, M. Hashimoto, H. Kazui, N. Hirono, and M. Matsui. Accelerated hippocampal atrophy in Alzheimer's disease with apolipoprotein E epsilon4 allele. *Ann. Neurol.*, 51 (2):209–214, 2002.

[51] J.D. Tapsoba, C. Kooperberg, A. Reiner, C.Y. Wang, and J.Y. Dai. Robust estimation for secondary trait association in casecontrol genetic studies. *Am. J. Epidemiol.*, 179 (10):1264–1272, 2014.

[52] J. Wang and S. Shete. Estimation of odds ratios of genetic variants for the secondary phenotypes associated with primary disease. *Genet. Epidemiol.*, 35:190–200, 2011.

[53] H. Zhu, Z. Khondker, Z. Lu, and J.G. Ibrahim. Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. *J. Am. Stat. Assoc.*, 109:977–990, 2014.

[54] M. Skup, H. Zhu, and H Zhang. Multiscale adaptive marginal analysis of longitudinal neuroimaging data with time-varying covariates. *Biometrics*, 68:1083–1092, 2012.

[55] Z. Xu, X. Shen, W. Pan, and Alzheimer's Disease Neuroimaging Initiative. Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. *PLoS ONE*, 9 (8):e102312, 2014.

[56] T. Ge, T.E. Nichols, D. Ghosh, E.C. Mormino, Smoller J.W., et al. A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. *NeuroImage*, 109 (1):505–514, 2015.

[57] K. Wang. Testing genetic association by regressing genotype over multiple phenotypes. *PLoS ONE*, 9 (9):e106918, 2014.

[58] P.F. OReilly, C.J. Hoggart, Y. Pomyen, F.C.F. Calboli, P. Elliott, et al. MultiPhen: Joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE*, 7:e34861, 2012.

[59] P. McCullagh. Regression models for ordinal data. *J. R. Stat. Soc. B*, 42 (2):109–142, 1980.

[60] W. Zhang, Z. Zhang, X. Li, and Q. Li. Fitting proportional odds model to case-control data with incorporating Hardy-Weinberg equilibrium. *Scientific Reports*, 5:17286, 2015.

[61] A.J. Trachtenberg, N. Filippini, K.P. Ebmeier, S.M. Smith, F. Karpe, and C.E. Mackay. The effects of APOE on the functional architecture of the resting brain. *Neuroimage*, 59:565–572, 2012.

[62] J.S. Damoiseaux, W.W. Seeley, J. Zhou, W.R. Shirer, G. Coppola, A. Karydas, et al. Gender modulates the apoe $\varepsilon 4$ effect in healthy older adults: convergent evidence from functional brain connectivity and spinal fluid tau levels. *J. Neurosci.*, 32:8254–8262, 2012.

[63] E. Walton, D. Geisler, J. Hass, J. Liu, J. Turner, et al. The impact of genome-wide supported schizophrenia risk variants in the neurogranin gene on brain structure and function. *PLoS One*, 8 (10):e76815, 2013.

[64] D.C. Glahn, A.M. Winkler, P. Kochunov, L. Almasy, R. Duggirala, et al. Genetic control over the resting brain. *Proc. Natl. Acad. Sci.*, 107:12231228, 2010.

[65] Y. Liu, T. Paajanen, E. Westman, L.O. Wahlund, A. Simmons, et al. Effect of APOE $\epsilon 4$ allele on cortical thicknesses and volumes: the AddNeuroMed study. *J. Alzheimers Dis.*, 21 (3):947–966, 2010.

[66] E.M. Tunbridge, S.M. Farrell, P.J. Harrison, and C.E. Mackay. Catechol-o-methyltransferase (COMT) influences the connectivity of the prefrontal cortex at rest. *Neuroimage*, 68:49–54, 2013.

[67] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52:1059–1069, 2010.

[68] J. Kim, J.R. Wozniak, B.A. Mueller, X. Shen, and W. Pan. Comparison of statistical tests for group differences in brain functional networks. *Neuroimage*, 101:681–694, 2014.

[69] S.E. Medland, N. Jahanshad, B.M. Neale, and P.M. Thompson. Whole-genome analyses of whole-brain data: working within an expanded search space. *Nat. Neurosci.*, 17 (6):791–800, 2014.

[70] P.M. Thompson, J.L. Stein, S.E. Medland, D.P. Hibar, A.A. Vasquez, and others. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.*, 8 (2):153–182, 2014.

[71] N. Jahanshad, P. Rajagopalan, X. Hua, D.P. Hibar, Nir. T.M., A.W. Toga, et al. Genome-wide scan of healthy human connectome discovers spon1 gene variant influencing dementia severity. *Proc. Natl. Acad. Sci.*, 110 (12):4768–4773, 2013.

[72] M.D. Greicius, G. Srivastava, A.L. Reiss, and V. Menon. Default mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional mri. *Proc. Natl. Acad. Sci.*, 101:4637–4642, 2004.

[73] O. Querbes, F. Aubry, J. Pariente, J.A. Lotterie, J.F. Démonet, V. Duret, M. Puel, et al. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain*, 132 (8):2036–2047, 2009.

[74] D. Liu, X. Lin, and D. Ghosh. Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. *Biometrics*, 63:1079–1088, 2007.

[75] W. Pan, J. Kim, Y. Zhang, X. Shen, and P. Wei. A powerful and adaptive association test for rare variants. *Genetics*, 197 (4):1081–1095, 2014.

[76] J. Kim and W. Pan. Highly adaptive tests for group differences in brain functional connectivity. *Neuroimage: Clinical*, 9:625–639, 2015.

[77] Rahul S. Desikan, F. Ségonne, B. Fischl, B.T. Quinn, B.C. Dickerson, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, 31 (3):968–980, 2006.

[78] J.S. Damoiseaux and M.D. Greicius. Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Struct. Funct.*, 213:525–533, 2009.

[79] M. Donix, A.C. Burggren, M. Scharf, K. Marschner, N.A. Suthana, et al. APOE associated hemispheric asymmetry of entorhinal cortical thickness in aging and Alzheimer's disease. *Psychiatry Res.*, 214 (3):212–220, 2013.

[80] L. Gutiérrez-Galve, M. Lehmann, N.Z. Hobbs, M.J. Clarkson, G.R. Ridgway, et al. Patterns of cortical thickness according to APOE genotype in Alzheimer's disease. *Dement. Geriatr. Cogn. Disord.*, 28 (5):476–485, 2009.

[81] L.Q. Uddin, A.M. Clare Kelly, B.B. Biswal, F.X. Castellanos, and M.P. Milham. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Hum. Brain Mapp.*, 30 (2):625–637, 2009.

[82] S. Passow, K. Specht, Tom C. Adamsen, M. Biermann, N. Brekke, A.R. Craven, et al. Default-mode network functional connectivity is closely related to metabolic activity. *Hum. Brain Mapp.*, 36 (6):2027–2038, 2015.

[83] P.M. Thompson, G. Tian, D.C. Glahn, N. Jahanshad, and T.E. Nichols. Genetics of the connectome. *NeuroImage*, 80:475–488, 2013.

[84] T.G. Lesnick, S. Papapetropoulos, D.C. Mash, J. Ffrench-Mullen, L. Shehadeh, et al. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet.*, 3 (6):e98, 2007.

[85] T. van Eimeren, O. Monchi, B. Ballanger, and A.P. Strafella. Dysfunction of the default mode network in Parkinson disease: a functional magnetic resonance imaging study. *Arch. Neurol.*, 66 (7):877–883, 2009.

[86] K. Venkova, A. Christov, Z. Kamaluddin, P. Kobalka, S. Siddiqui, and K. Hensley. Semaphorin 3A signaling through Neuropilin-1 is an early trigger for distal axonopathy in the SOD1G93A mouse model of amyotrophic lateral sclerosis. *J. Neuropathol. Exp. Neurol.*, 73 (7):702–713, 2014.

[87] R.J. Pruim, R.P. Welch, S. Sanna, T.M. Teslovich, P.S. Chines, T.P. Gliedt, et al. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, 26:2336–2337, 2015.

[88] L. Morris, S. Veeriah, and T. Chan. Genetic determinants at the interface of cancer and neurodegenerative disease. *Oncogene*, 29 (24):3453–3464, 2010.

[89] Y. Hirata, C.C. Zai, R.P. Souza, J.A. Lieberman, H.Y. Meltzer, and J.L. Kennedy. Association study of GRIK1 gene polymorphisms in schizophrenia: case-control and family-based studies. *Hum. Psychopharmacol.*, 27 (4):345–51, 2012.

[90] H. Shibata, A. Joo, Y. Fujii, A. Tani, et al. Association study of polymorphisms in the Glur5 kainate receptor gene (GRIK1) with schizophrenia. *Psychiatr. Genet.*, 11 (3):139–144, 2001.

[91] Alzheimer's Association. Alzheimer's Disease Facts and Figures. *Alzheimer's and Dementia*, 11:332–384, 2015.

[92] Alzheimer's Association. Changing the trajectory of alzheimer's disease: How a treatment by 2025 saves lives and dollars. *Alzheimer's and Dementia*, 2015.

[93] H. Marei, A. Althani, M. El Zowalaty, M.A. Albanna, et al. Common and rare variants associated with alzheimer's disease. *J. Cell Physiol.*, 2015.

[94] A.J. Saykin, L. Shen, X. Yao, S. Kim, K. Nho, et al. Genetic studies of quantitative mci and ad phenotypes in adni: Progress, opportunities, and plans. *Alzheimer's and Dementia*, 11:792–814, 2015.

[95] L. Jones, P.A. Holmans, M.L. Hamshere, D. Harold, V. Moskvina, D. Ivanov, et al. Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimers disease. *PLoS ONE*, 5:e13950, 2010.

[96] M.G. Hong, C.A. Reynolds, A.L. Feldman, M. Kallin, et al. Genome-wide and gene-based association implicates FRMD6 in alzheimer disease. *Hum. Mutat.*, 33:521–529, 2012.

[97] R. Sherva, Y. Tripodis, D.A. Bennett, L.B. Chibnik, P.K. Crane, et al. Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimer's Dement.*, 10:45–52, 2014.

[98] B. Metin, R.M. Krebs, J.R. Wiersema, T. Verguts, R. Gasthuys, et al. Dysfunctional modulation of default mode network activity in attention-deficit/hyperactivity disorder. *Abnorm. Psychol.*, 124 (1):208–214, 2015.

[99] Y. He, Z. Chen, G.L. Gong, and A. Evans. Neuronal networks in alzheimer's disease. *Neuroscientist*, 15:333–350, 2009.

[100] D.T. Jones, M.M. Machulda, P. Vemuri, E.M. McDade, G. Zeng, M.L. Senjem, et al. Age-related changes in the default mode network are more advanced in Alzheimers disease. *Neurology*, 77 (16):1524–1531, 2011.

[101] M. Balthazar, M. Weiler, B. Campos, T. Rezende, B. Damasceno, and F. Cendes. Alzheimer as a default mode network disease: A grey matter, functional and structural connectivity study. *Neurology*, 83 (10):P6.324, 2014.

[102] Y. Wang, A. Liu, J.L. Mills, M. Boehnke, A. F. Wilson, J.E. Bailey-Wilson, et al. Pleiotropy analysis of quantitative traits at gene level by multivariate functional linear models. *Genet. Epidemiol.*, 39 (4):259–75, 2015.

[103] C.S. Tang and M.A.R. Ferreira. A gene-based test of association using canonical correlation analysis. *Bioinformatics*, 28 (6):845–850, 2012.

[104] H. Aschard, B. Vilhjalmsson, C. Wu, N. Greliche, P.E. Morange, et al. Maximizing the power in principal components analysis of correlated phenotypes. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, 94 (5):662–676, 2014.

[105] S. Mukherjee, S. Kim, V.K. Ramanan, LE. Gibbons, et al. Gene-based GWAS and biological pathway analysis of the resilience of executive functioning. *Brain Imaging Behav.*, 8:110–118, 2014.

[106] W. Pan. Relationship between genomic distance-based regression and kernel machine regression for multi-marker association testing. *Genet. Epidemiol.*, 35 (4):211–216, 2011.

[107] B.H. McArdle and M.J. Anderson. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology*, 82:290–297, 2001.

[108] J.Y. Tzeng, D. Zhang, M. Pongpanich, C. o Smith, et al. Studying gene and gene-environment effects of uncommon and common variants on continuous traits: a marker-set approach using gene-trait similarity regression. *Am. J. Hum. Genet.*, 89:277–288, 2011.

[109] J. Wessel and N.J. Schork. Generalized genomic distance-based regression methodology for multilocus association analysis. *Am. J. Hum. Genet.*, 79:792–806, 2006.

[110] M.A. Zapala and N.J. Schork. Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. *Front. Genet.*, 3:190, 2012.

[111] D.J. Schaid, S.K. McDonnell, S.J. Hebbring, J.M. Cunningham, et al. Nonparametric tests of association of multiple genes with human disease. *Am. J. Hum. Genet.*, 76:780–793, 2005.

[112] R.F. Haase. *Multivariate General Linear Models*. SAGE Publications, 2011.

[113] K.E. Muller and B.L. Peterson. Practical methods for computing power in testing the multivariate general linear hypothesis. *Comput. Stat. Data An.*, 2:143–158, 1984.

[114] W. Pan, I. Kwak, and P. Wei. A powerful pathway-based adaptive test for genetic association with common or rare variants. *Am. J. Hum. Genet.*, 97:86–98, 2015.

[115] S.A. Meda, G. RuaÓo, A. Windemuth, K. O'Neil, C. Berwise, S.M. Dunn, et al. Multivariate analysis reveals genetic associations of the resting default mode network in psychotic bipolar disorder and schizophrenia. *Proc. Natl. Acad. Sci.*, 111 (19):E2066–2075, 2014.

[116] R.L. Buckner, J.R. Andrews-Hanna, and D.L. Schacter. The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.*, 1124:1–38, 2008.

[117] T. Hamatani, T. Daikoku, H. Wang, H. Matsumoto, M.G. Carter, et al. Global gene expression analysis identifies molecular pathways distinguishing blastocyst dormancy and activation. *Proc. Natl. Acad. Sci.*, 101 (28):10326–10331, 2004.

[118] R.J. Anney, Lasky-Su J., C. O'Dúshláine, E. Kenny, B.M. Neale, et al. Conduct disorder and adhd: evaluation of conduct problems as a categorical and quantitative trait in the international multicentre adhd genetics study. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, 147B (8):1369–1378, 2008.

[119] J.F. Schmouth, M. Castellarin, S. Laprise, K.G. Banks, R.J. Bonaguro, et al. Non-coding-regulatory regions of human brain genes delineated by bacterial artificial chromosome knock-in mice. *BMC biol.*, 11:106, 2013.

[120] G. Liu, L. Yaoc, J. Liu, Y. Jiang, et al. Cardiovascular disease contributes to alzheimer's disease: evidence from large-scale genome-wide association studies. *Neurobiol. of Aging*, 35 (4):786–792, 2014.

[121] M.I. Kamboh, F.Y. Demirci, X. Wang, R.L. Minster, et al. Genome-wide association study of alzheimer's disease. *Transl. Psychiatry*, 15 (2):e117, 2012.

[122] S. Seshadri, A.L. Fitzpatrick, M. Arfan Ikram, A.L. DeStefano, V. Gudnason, et al. Genome-wide analysis of genetic loci associated with Alzheimer's disease. *JAMA*, 303 (18):1832–1840, 2010.

[123] EY. Liu, M. Li, W. Wang, and Y. Li. Mach-admix: Genotype imputation for admixed populations. *Genet. Epidemiol.*, 37 (1):25–37, 2013.

[124] C.H. Chen, Q. Peng, A.J. Schork, M.T. Lo, C.C. Fan, et al. Large-scale genomics unveil polygenic architecture of human cortical surface area. *Nat. Commun.*, 6:7549, 2015.

[125] L. Wang, J. Zhou, and A. Qu. Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics*, 68 (2):353–360, 2012.

[126] W.J. Fu. Penalized estimating equationsn. *Biometrics*, 59 (1):126–132, 2003.

Appendix A

Simulations for GEE-aSPUpath

A.1 Simulation set-up

We conducted a small simulation study to demonstrate the performance of the GEE-aSPUpath test. The simulated data mimicked the ADNI-1 dataset. The phenotype data were simulated based on the grey matter volumes in the 12 ROIs corresponding to the default mode network (DMN). For covariates, we included gender, education, handedness, age and ICV as available in the ADNI-1 dataset. We used a KEGG pathway `hsa00410` containing 20 genes with 592 SNPs in total.

We explored two factors that might influence testing power: 1) the overall effect size of the pathway and 2) varying gene-level and trait-level association patterns. In simulation set-up 1, we varied only pathway effect sizes. Using the ADNI-1 data, first, we estimated the marginal effect of each SNP j on an individual trait t by estimating the regression coefficients (i.e. w_{jt}), and estimated each covariate effect (q) on each trait (i.e. ψ_{qt}). The sample covariance matrix of the multiple traits (Σ) was evaluated. Denote the mean vector of the 12 traits W_0, and the estimated regression coefficient matrices $W = (w_{jt})$ and $\Psi = (\psi_{qt})$.

Given a pathway G with $|G|$ genes, the genotype scores for the SNPs in the pathway for subject i are $x_i = (x'_{i,1}, ..., x'_{i,|G|})'$ with gene g including h_g SNPs, $x_{i,g} = (x_{i,g,1}, ..., x_{i,g,h_g})'$. To maintain the original correlation structures among the genotype scores x_i and the five covariates $zi = (z_{i1}, ..., z_{i5})'$, we used every pair of (x_i, z_i) from the ADNI-1 data in each simulation. The multiple traits for subject i were generated

from a multivariate normal distribution:

$$Y_i \sim \mathcal{MN}(W_0 + \phi \cdot W' x_i + \Psi' z_i, \ \ \Sigma).$$

Here a scaling factor ϕ was used to control the effect sizes of the pathway: with $\phi = 0$, there was no association and Type I error rates were evaluated; as ϕ increased, the association strengths of the pathway with the multiple traits increased and power was evaluated.

In simulation set-up 2, we considered the presence of non-associated SNP-trait pairs, which is expected to be more realistic than set-up 1. Out of the 20 genes in the pathway `hsa00410`, 10 genes were defined as causal; in each causal gene, we randomly selected two-thirds of the SNPs as causal and the rest as null SNPs; all the SNPs in a non-causal gene were null. We also restricted each causal SNP to be associated with only 8 or 9 traits out of the total of 12 traits. We designated a 0 regression coefficient for each non-associated SNP-trait pair; otherwise, the same regression coefficients were used for others. As before, (x_i, z_i) pairs were sampled from the ADNI-1 data, and Y_i was generated from the multivariate normal distribution.

Throughout simulations, 1000 replicates were used and the tests were conducted at the significance level $\alpha = 0.05$. For the GEE-aSPUset test, we used $\gamma_1, \gamma_2 \in \{1, ..., 8\}$; for the GEE-aSPUpath test, $\gamma_1 \in \{1, ..., 8\}$ and $\gamma_2, , \gamma_3 \in \{1, 2, 4, 8\}$. We used $B = 1000$ permutations for each test.

A.2 Type I error and power

The empirical Type I error rates (with $\phi = 0$) were well controlled by both GEE-aSPUpath and GEE-aSPUset tests (Table A.1). As the effect sizes (controlled by $\phi > 0$) of the pathway increased, the power of both GEE-aSPUpath and GEE-aSPUset tests increased; the GEE-aSPUset test performed better, since in set-up 1 all SNPs in the pathway were causal, for which the adaptiveness of the GEE-aSPUpath test to the genes was useless. Note that set-up 1 was not realistic with all SNP-trait pairs being associated (for $\phi > 0$).

For simulation set-up 2 (Table A.2), perhaps due to the varying association patterns at both the gene-level and trait-level, the GEE-aSPUpath test was slightly more powerful than the GEE-aSPUset test.

Table A.1: Type I errors ($\phi = 0$) and power ($\phi \neq 0$) under varying overall pathway effect size.

ϕ	GEE-aSPUpath	GEE-aSPUset
0	0.050	0.0495
0.02	0.073	0.100
0.04	0.110	0.332
0.06	0.273	0.847
0.08	0.654	0.998
0.10	0.951	1.000

Table A.2: Power under varying pathway effect size and sparsity of associations.

ϕ	GEE-aSPUpath	GEE-aSPUset
0.05	0.084	0.066
0.08	0.142	0.130
0.10	0.236	0.208
0.15	0.663	0.604
0.18	0.894	0.874
0.20	0.980	0.961