RESEARCH ARTICLE

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

Junghi Kim[1] | Wei Pan[1] | for the Alzheimer's Disease Neuroimaging Initiative[2]

[1]Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, United States of America

[2]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how to apply/ADNI Acknowledgement List.pdf.

Correspondence
Wei Pan, Division of Biostatistics, MMC 303, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA.
Email: weip@biostat.umn.edu

ABSTRACT

There has been increasing interest in developing more powerful and flexible statistical tests to detect genetic associations with multiple traits, as arising from neuroimaging genetic studies. Most of existing methods treat a single trait or multiple traits as response while treating an SNP as a predictor coded under an additive inheritance mode. In this paper, we follow an earlier approach in treating 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 it is 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, 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 Alzheimer's Disease Neuroimaging Initiative (ADNI) data, in which either structural MRI driven phenotypes or resting-state functional MRI (rs-fMRI) derived brain functional connectivity measures were used as phenotypes. The applications led to the identification of several top SNPs of biological interest. Furthermore, simulation studies showed competitive performance of the new method, especially for $p > n$.

KEYWORDS
ADNI, aSPU, default mode network (DMN), functional connectivity, GWAS, high dimensional phenotypes, MRI, rs-fMRI

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) (Bigos, Hariri, & Weinberger, 2016; Shen et al., 2014). 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 (Ferreira, & Purcell, 2009; He, Avery, & Lin, 2013; O'Reilly et al., 2012; Schifano, Li, Christiani, & Lin, 2013; van der Sluis, Posthuma, & Dolan, 2013; Wang, 2014; Zhang, Xu, Shen, & Pan, 2014), association analyses for multiple phenotypes are challenging, because a uniformly most powerful test does not exist. A key issue in multitrait analysis is how to maximize the statistical power in the presence of many nonassociated 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 (van der Sluis et al., 2013), or using principal components analysis (PCA), principal components of heritability (PCH) or related methods for dimension reduction (Du et al., 2016; Ferreira, & Purcell, 2009; Klei et al., 2008; Lin et al., 2012; Sun et al., 2015; Vounoua et al., 2010; Wang, & Abbott, 2007). 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 (Lin, & Tang, 2011; Shen et al., 2010) and variance component tests (He et al., 2013). 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. (2014) 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 up-weights the traits more highly associated with the SNP, in which way the test's power remains high even in the presence of many nonassociated 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. (2014) 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 (2014), the use of the additive inheritance model may lead to loss of power when the assumption is violated. We note that most existing methods are not applicable to cases with $p > n$.

In this paper, 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. The POM (McCullagh, 1980) assumes that there is a continuous unmeasured latent variable whose values determine the observed ordinal values (i.e., genotype scores), and the cut-off points of the latent variable for the ordinal values appear as an intercept term in cumulative logits of the ordinal variable. The model assumes identical log-odds ratios across cumulative logits, but the intercept depends on the category, which allows a nonlinear relationship between the genotype and the phenotype. 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 (O'Reilly et al., 2012; Wang 2014; Zhang, Zhang, Li, & Li, 2015), 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 its variants, or sparse regression (Damoiseaux et al., 2012; Du et al., 2016; Lin et al., 2012; Sun et al., 2015; Trachtenberg et al., 2012; Vounoua et al., 2010; Wang, Sha, & Zhang, 2016). 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 (Zhang et al., 2014).

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, and volumes (Du et al., 2016; Shen et al., 2010; Walton et al., 2013) 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 (Damoiseaux et al., 2012; Glahn et al., 2010; Liu et al., 2010; Trachtenberg et al., 2012; Tunbridge, Farrell, Harrison, & Mackay, 2013). 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 (Kim, Wozniak, Mueller, Shen, & Pan, 2014; Rubinov, & Sporns, 2010); 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 (Medland, Jahanshad, Neale, & Thompson, 2014; Thompson et al., 2014). Several studies conducted GWAS for brain connectivity analyses, but used only single connectivity (Jahanshad et al. 2013; Medland et al., 2014), 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 (Damoiseaux et al., 2012; Greicius, Srivastava, Reiss, & Menon, 2004; Trachtenberg et al., 2012). In particular, cortical gray matter (GM) thicknesses from DMN ROIs

were employed for its capability of detecting preclinical Alzheimer's disease (Querbes et al., 2009). In addition, we considered functional connectivity in DMN as multiple phenotypes, which are useful but underutilized 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 (Zhang et al., 2014) and potential power gains when the genetic inheritance mode was nonadditive but dominant.

In the following, we introduce the new adaptive test in a POM, then the new method and other methods are compared with applications to the ADNI data and simulated data. We end with a short summary of the conclusions and future directions.

2 | METHODS

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 $X_i = (x_{i1}, \dots, x_{ip})$ and l covariates $Z_i = (z_{i1}, \dots, z_{il})$ for $i = 1, \dots, n$. Define $r_{ij} = Pr(Y_i \leq j | X_i, Z_i)$ for $j = 0, 1, 2$.

First we describe the POM, which is widely used for ordinal response data (McCullagh, 1980; O'Reilly et al., 2012; Wang et al. 2014). Define two sets of regression coefficient vectors: $\beta = (\beta_1, \dots, \beta_p)'$ and $\delta = (\delta_1, \dots, \delta_l)'$, and a vector of intercepts $\alpha = (\alpha_0, \dots, \alpha_{J-2})'$. The cumulative logit model becomes

$$\text{logit}[Pr(Y_i \leq j)] = \alpha_j + Z_i\delta + X_i\beta \quad j = 0, 1. \quad (1)$$

This model assumes that Z or X have identical effects across 2 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 (1) can be derived based on the multinomial distribution for the categorical variable Y_i. McCullagh (1980) reparameterized the likelihood in terms of a cumulative probability r_{ij}. The $(J + l + p - 1)$ dimensional score vector for POM in equation (1) is a gradient of the log likelihood with respect to $\theta = (\alpha', \delta', \beta')'$: $U_\theta = (U_\alpha', U_\delta', U_\beta')'$. Following McCullagh's (1980) approach, we derive a closed form for each component of the score, U_α, U_δ, and U_β, as shown in Appendix A. We estimate the covariance matrix of the score vector $Cov(U_\theta)$ based on the observed Fisher information matrix as shown in Appendix B. The covariance matrix can be partitioned according to the parameter components (α', δ') 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 $H_0 : \beta = (\beta_1, \dots, \beta_p)' = 0$ using the score vector

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

where $\hat{r}_{ij} = exp(\hat{\alpha}_j + Z_i\hat{\delta})/[1 + exp(\hat{\alpha}_j + Z_i\hat{\delta})]$ for $j = 0$ or 1 is from the fitted null model of equation (1); $\hat{\alpha}_j$ and $\hat{\delta}$ 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 $\hat{\alpha}_j$ and $\hat{\delta}$ 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 (Schifano et al., 2013; Wang et al. 2014). 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.

2.2 | An adaptive test

Suppose U_k is the kth component of the score vector $U = (U_1, \dots, U_p)'$. 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 (Lin, & Tang, 2011) 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 (e.g., kernel machine regression), respectively (Liu, Lin, & Ghosh, 2007; Pan, Kim, Zhang, Shen, & Wei, 2014). 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 SPU(∞) = $\max_{k=1}^{p}|U_k|$, 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 $\gamma \geq 1$; $\Gamma = \{1, 2, \dots, 8, \infty\}$ was used for its good performance in all numerical studies.

The optimal range of γ's in Γ depends on the true but unknown genetic architecture; for practical use, Pan et al. (2014) and Kim, Zhang, & Pan (2016) provided a general guidance. Suppose a set of candidate γs are given in $\Gamma = \{1, 2, \ldots, C_1, \infty\}$. We can define C_1 such that the SPU(C_1) test gives a P-value close to that of SPU(∞). For a larger number of phenotypes, a larger value of C_1 may be required. In general, if the association pattern is believed to be sparse (i.e., with only few associated SNP-trait pairs), then using larger γs may give higher power; vice versa, SPU(γ) with a smaller γ is more powerful for a higher proportion of associated SNP-trait pairs. When the traits are expected to be associated with the SNP in opposite directions, using only even integers in Γ may be most powerful; on the other hand, if the most associations between SNP-trait pairs are in one direction, only odd integers are needed in Γ; if there is no prior knowledge, as the default, both even and odd integers should be used.

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 $b = 1, \ldots, B$. Then the null statistics SPU(γ)$^{(b)}$ are obtained for each b. The P-value of each SPU(γ) is calculated as $P_{\text{SPU}(\gamma)} = \left[\sum_{b=1}^{B} I(|\text{SPU}(\gamma)^{(b)}| \geq |\text{SPU}(\gamma)|) + 1 \right]/(B + 1)$, 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 $P_{\text{aSPU}} = \left[\sum_{b=1}^{B} I(\text{aSPU}^{(b)} \leq \text{aSPU}) + 1 \right]/(B + 1)$ where $\text{aSPU}^{(b)} = \min_{\gamma \in \Gamma} P_{\gamma}^{(b)}$ and $P_{\gamma}^{(b)} = \left[\sum_{b_1 \neq b} I(|\text{SPU}(\gamma)^{(b_1)}| \geq |\text{SPU}(\gamma)^{(b)}|) + 1 \right]/B$.

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 $\{1, 2, \ldots, n\}$ is permuted to $\{\sigma(1), \sigma(2), \ldots, \sigma(n)\}$; replace X_i in equation (2) with $X_{\sigma(i)}$. With the null score $U^{(b)}$ obtained from each permutation b, the null statistics SPU(γ)$^{(b)}$ are computed for each γ. The P-values of each SPU(γ) 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.

2.3 | A doubly adaptive test

Suppose we consider p connectivity phenotypes as a brain network. Differences in brain networks can result from differences in genotype. Due to a large number of parameters in estimating a network, often a penalized method is applied to strike a good bias-variance trade-off in the resulting estimate (Lin et al., 2012; Vounoua et al., 2010). However, choosing the regularization parameter to minimize the estimation or prediction error does not necessarily lead to high power in testing; an optimal procedure for estimating networks may no longer be optimal for hypothesis testing as illustrated previously (Kim, Wozniak, Mueller, & Pan, 2015a; Kim, & Pan, 2015b). Accordingly, we have to choose the regularization parameter to maximize testing power in the current context.

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 rewritten as

$$\text{logit}[Pr(Y_i \leq j)] = \alpha_j + Z_i \delta + X_i(t) \cdot \beta, \quad j = 0, 1,$$

$$U(t) = \sum_{i=1}^{n} \sum_{j=0}^{J-2} (1 - \hat{r}_{i(j-1)} - \hat{r}_{ij}) \cdot I(Y_i = j) \cdot X_i(t).$$

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

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

$$\text{aSPU}(\gamma) = \min_t P_{\text{SPU}(t, \gamma)},$$

$$\text{daSPU} = \min_\gamma P_{\text{aSPU}(\gamma)},$$

where $U_k(t)$ is the kth element of $U(t)$. P-values of SPU(t, γ) and aSPU(γ), daSPU tests can be obtained similarly as before, based on the same set of simulated or permuted null scores $U^{(b)}$ for $b = 1, \ldots, B$. 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 SPU(t, γ)$^{(b)}$ for each γ's and t's.

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

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

Step 4. Based on the above null statistics, the P-values of SPU(t, γ), aSPU(γ), daSPU tests are obtained:
$P_{\text{SPU}(t,\gamma)} = \left[\sum_{b=1}^{B} I(|\text{SPU}(t, \gamma)^{(b)}| \geq |\text{SPU}(t, \gamma)|) + 1 \right]/(B + 1)$,

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

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

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., 2014)

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

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

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. (2012) proposed the likelihood ratio test, while Wang et al. (2014) derived the score test for POM. Both approaches assume a large samples size with $n > p$, 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 (Pan et al., 2014; Zhang et al., 2014).

We will compare the performance of the proposed method to that of GEE-based tests (Zhang et al., 2014), POM-based tests (Wang et al. 2014, O'Reilly et al., 2012), TATES (van der Sluis et al., 2013), MANOVA and MDMR (with the Euclidean distance metric) (McArdle, & Anderson, 2001) through real data analysis and simulation studies.

3 | RESULTS

3.1 | Real data example

3.1.1 | ADNI data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit 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 coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the United States 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 1,500 adults, ages 55–90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

3.1.2 | 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 (Zhang et al., 2014). We considered some candidate SNPs: rs429358, rs2075650, rs7526034, rs10932886, rs7647307, rs7610017, rs4692256, and rs6463843, which were shown to be strongly associated with some quantitative imaging traits (Shen et al., 2010). From the ADNI-1 baseline scans, the cortical thicknesses for 68 ROIs were extracted based on the Desikan-Killany atlas (Desikan et al., 2006). The sample size was $n = 638$ with 145 ADs, 182 normal controls (CNs) and 311 subjects with minor cognitive impairment (MCIs).

TABLE 1 *p*-values for association testing between DMN cortical thickness and each candidate SNP

rs ID	chr	MAF	POM Score	perm-aSPU	GEE Score	perm-aSPU	O'Reilly MultiPhen	van der Sluis TATES	MANOVA	McArdle MDMR
rs429358	19	0.30	5.17e-05	2.00e-08	2.40e-07	2.90e-08	3.51e-05	9.71e-07	8.16e-05	9.99e-04
rs2075650	19	0.25	1.35e-05	9.00e-07	1.52e-06	4.90e-07	8.46e-06	4.36e-06	1.91e-05	1.99e-06
rs7526034	1	0.12	2.53e-02	1.30e-03	3.50e-02	6.00e-04	2.30e-02	3.24e-04	7.58e-03	1.99e-03
rs10932886	2	0.32	2.89e-03	2.00e-03	4.26e-03	5.70e-03	2.14e-03	5.43e-03	4.74e-03	3.99e-03
rs7647307	3	0.44	7.34e-03	1.52e-02	2.10e-03	1.26e-02	4.45e-03	2.93e-03	8.09e-03	1.99e-03
rs7610017	3	0.04	5.65e-01	2.16e-01	6.37e-01	2.49e-01	5.76e-01	4.17e-01	6.15e-01	2.46e-01
rs4692256	4	0.46	8.42e-02	1.04e-01	1.83e-01	1.29e-01	8.53e-02	8.45e-02	8.93e-02	6.39e-02
rs6463843	7	0.47	1.04e-02	1.20e-03	6.61e-03	7.00e-04	6.57e-03	1.06e-04	8.88e-03	4.99e-04

TABLE 2 *p*-values for association testing between 68 regions' cortical thickness and each candidate SNP

rs ID	chr	MAF	POM Score	perm-aSPU	GEE Score	perm-aSPU	O'Reilly MultiPhen	van der Sluis TATES	MANOVA	McArdle MDMR
rs429358	19	0.30	6.15e-04	3.60e-06	1.55e-06	1.35e-05	5.18e-05	1.76e-06	4.38e-04	1.99e-06
rs2075650	19	0.25	4.99e-02	3.00e-05	9.07e-03	3.60e-06	2.09e-02	7.58e-06	4.22e-02	1.98e-06
rs7526034	1	0.12	5.11e-01	2.00e-04	1.64e-03	2.00e-04	3.97e-04	2.39e-05	6.30e-04	3.99e-05
rs10932886	2	0.32	1.73e-02	2.17e-02	2.05e-02	2.85e-02	6.04e-03	1.28e-02	1.52e-02	2.09e-02
rs7647307	3	0.44	1.07e-02	1.48e-02	5.57e-02	1.16e-02	4.12e-02	3.16e-03	1.09e-01	3.99e-03
rs7610017	3	0.04	5.51e-01	2.67e-01	5.94e-01	2.82e-01	4.86e-01	5.26e-01	6.16e-01	2.46e-01
rs4692256	4	0.46	1.29e-01	2.11e-02	1.15e-01	1.37e-01	5.97e-02	8.94e-02	1.27e-01	3.29e-02
rs6463843	7	0.47	3.85e-01	1.10e-03	1.27e-01	2.00e-04	2.46e-01	5.66e-04	3.78e-01	2.99e-04

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 (Damoiseaux & Greicius, 2009; Greicius et al., 2004). 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 1 and 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, respectively. 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 (Donix et al., 2013; Gutiérrez-Galve et al., 2009; Liu et al., 2010).

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

Neuroimaging traits may be in the number of hundreds to thousands while only a portion of them are expected to be associated with an SNP. We conducted a GWAS scan with

high dimensional functional connectivity traits. The genotype data and rs-fMRI data were obtained from ADNI-2. For genotype data, we included all SNPs with a minor allele frequency (MAF) ≥ 0.05, genotyping rate > 90%, and surviving the Hardy-Weinberg equilibrium test with a *p*-value > 0.001. After all rounds of quality control, 578,175 SNPs remained. In rs-fMRI data, 116 brain regions were predefined as seed regions, and at each region, neuronal activity was measured in BOLD time series; pairwise correlations were computed between the BOLD time courses of various regions throughout the brain, and used as functional connectivity (Kim et al., 2014; Kim, & Pan, 2015b; Thompson, Tian, Glahn, Jahanshad, & Nichols, 2013). We considered two sets of high dimensional phenotypes. The first set was functional connectivity related to the default mode network and another was a brain-wide functional connectivity. $n = 134$ subjects were included, consisting of 24 ADs, 22 late MCIs (LMCIs), 44 early MCIs (EMCIs), 20 subjects with symptoms of memory loss (SMCs) and 24 CNs.

First, we identified 18 brain regions related to the default mode network (DMN) and defined them 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 (Greicius et al., 2004; Passow et al., 2015; Uddin et al. 2009). Given a set of nodes, functional
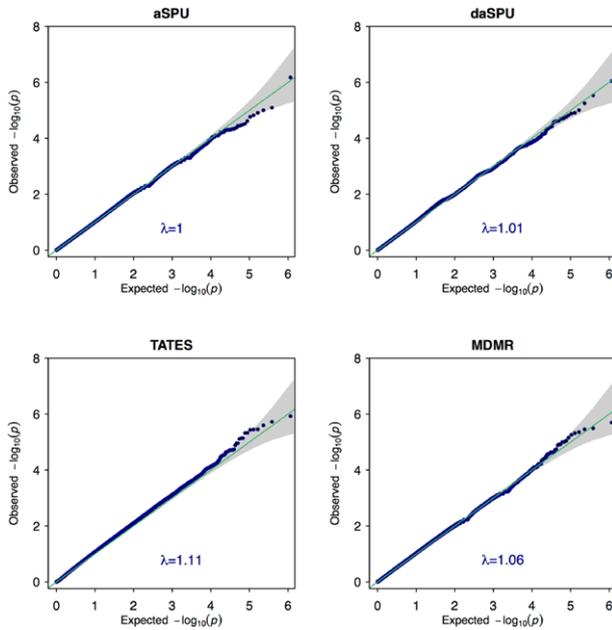
FIGURE 1 QQ plots from GWAS for function connectivity in DMN

connectivity between every pair of 18 nodes was calculated with the Pearson's correlation; a total of $p = 18 \times (18 - 1)/2$ functional connectivity was estimated, and Fisher's z-transformation was applied to each connection. The permutation based POM-aSPU, POM-daSPU, TATES, and MDMR were applied to each of 578,175 SNPs to test its association with the DMN functional network after adjusting for age and gender. We applied the doubly adaptive test (daSPU) for testing associations with regularized networks, by thresholding the empirical correlation matrix with candidate thresholds $t \in \{0, 0.1, 0.2, \ldots, 0.9\}$. After thresholding, Fisher's z-transformation was applied to each connection. Other methods were not considered here because they cannot handle the case with $p > n$. The QQ plots from the GWAS scan with each method are illustrated in Figure 1: all the inflation factors (λ) were reasonable as shown in each QQ plot. Figure 2 shows the Manhattan plots from the GWAS scan with the POM-aSPU, POM-daSPU, TATES, and MDMR, respectively. The POM-aSPU test succeeded in combining the significant associations identified by the individual SPU(γ) tests as shown in the Manhattan plots. Although none of the SNPs was significant at the genome-wide significance level, it was perhaps due to the small sample size.

The p-values for a few top significant SNPs identified by each method are reported in LocusZoom plots (Pruim et al., 2010). Figures 3, 4, 5, and 6 illustrate the results for POM-aSPU, POM-daSPU TATES, and MDMR, respectively. According to the POM-aSPU test (Fig. 3), 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 (Lesnick et al., 2007), and also related to dysfunction in DMN

(van Eimeren, Monchi, Ballanger, & Strafella, 2009). Gene NRP1 (rs2804498) has been implicated in Alzheimer disease, combined with another gene SEMA3A (Venkova et al., 2014). Among top significant SNPs identified by POM-daSPU test (Fig. 4), SNPs rs1412096 in gene PTPRD and rs7276462 in gene GRIK1 were additionally identified, which were discussed as possible triggers to AD (Hirata et al., 2012; Morris, Veeriah, & Chan, 2010; Shibata et al., 2001). The top SNPs identified by TATES (Fig. 5) were rs12082932 (gene RAB3GAP2), rs17108201 (gene ADAM), and rs9973183. Gene RAB3GAP2 including rs12082932 is associated with autophagy, whose modulation has been shown to affect neurodegeneration such as Alzheimer and Huntington disease (Spang et al., 2014). Brocker, Vasiliou, & Nebert (2009) discussed that gene ADAM including rs17108201 represents promising drug targets for the prevention and management of a number of human diseases. Many of top significant SNPs identified by MDMR were overlapped with top SNPs that POM-aSPU/POM-daSPU discovered (Fig. 6). In Table 3, we combined the results for those SNPs. Table 3 shows the $\hat{\gamma}$ values by which the SPU($\hat{\gamma}$) gave the minimum P-value among the SPU(γ) tests applied for POM-aSPU test and also presents the $(\hat{\gamma}, \hat{t})$ values at which the statistics for daSPU was defined. Among the SPU(γ) tests applied with $\gamma \in \{1, \ldots, 8, \infty\}$, 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.

For a higher dimensional phenotype, we applied the POM-aSPU test to the brain-wide functional connectivity network. All 116 brain regions were included as nodes, and a total of $p = 116 \times (116 - 1)/2$ pairwise correlations were estimated for the brain-wide functional connectivity as phenotypes from $n = 134$ samples. 578,175 SNPs was tested after adjusting age and gender. The other methods were excluded due to their inapplicability or slowness for such a case with $n << p$. In particular, TATES requires calculating the eigen values for a p by p matrix, which would be time-consuming with $p = 6670$. Figure 7 presents a QQ plot and a Manhattan plot from association testing for the whole brain functional connectivity. No SNP could pass the significance thresholds of 5e-08. Table 4 and Figure 8 illustrate the top four most significant SNPs; we observed that the SNPs in gene SV2C had some associations with the whole brain network. Gene SV2C on chromosome 5 is reported to be involved in Parkinson's disease pathogenesis in a previous study (Hill-Burns et al., 2013).

3.2 | Simulations

We carried out simulation studies to further investigate the performance of the proposed method as compared with
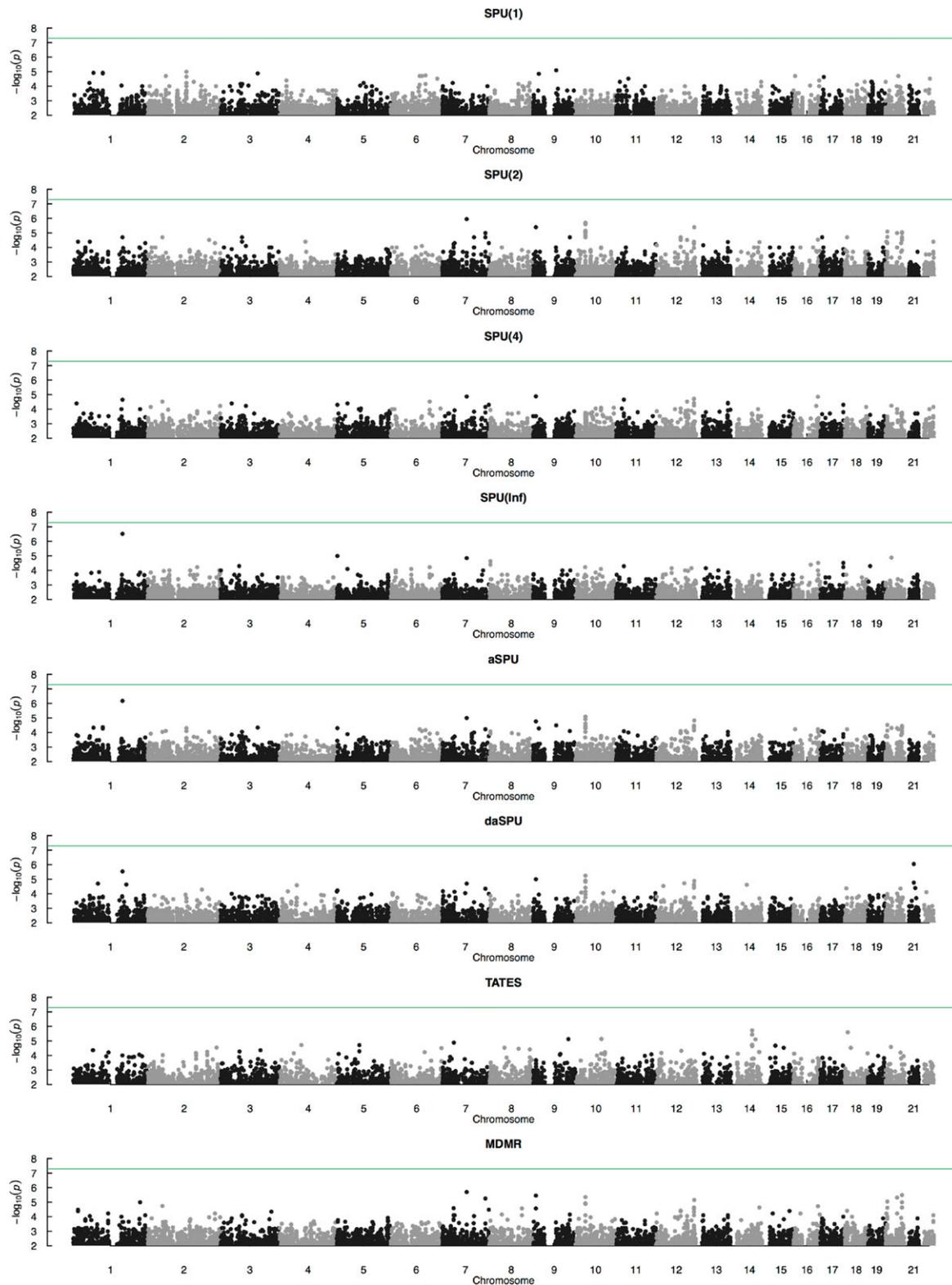
FIGURE 2 Manhattan plots from GWAS for function connectivity in DMN

GEE-based tests (Zhang et al., 2014), POM-based tests (O'Reilly et al., 2012; Wang et al. 2014), TATES (van der Sluis et al., 2013), MANOVA, and MDMR (with the Euclidean distance metric) (McArdle, & Anderson, 2001). By default, we considered a sample size $n = 1,000$ at each simulated dataset. Empirical Type I error rates and power were evaluated based on 1,000 replicates at significance level $\alpha = 0.05$ for each simulation scenario. For SPU, aSPU, and MDMR tests, $B = 1,000$ simulations or permutations were used to estimate their P-values. Two factors were considered in the simulation studies: genetic effect size and proportion of associated phenotypes.
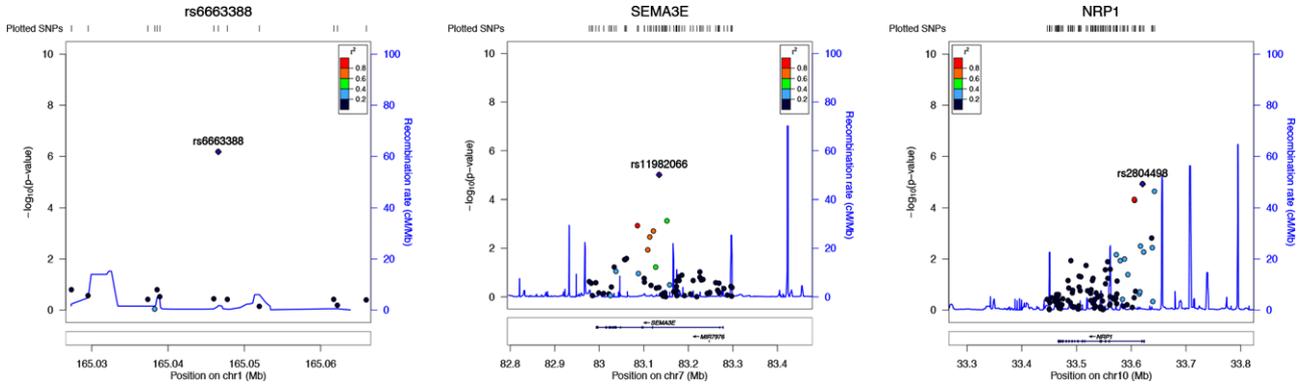
FIGURE 3 LocusZoom for top significant SNPs determined by POM-aSPU for functional connectivity in DMN
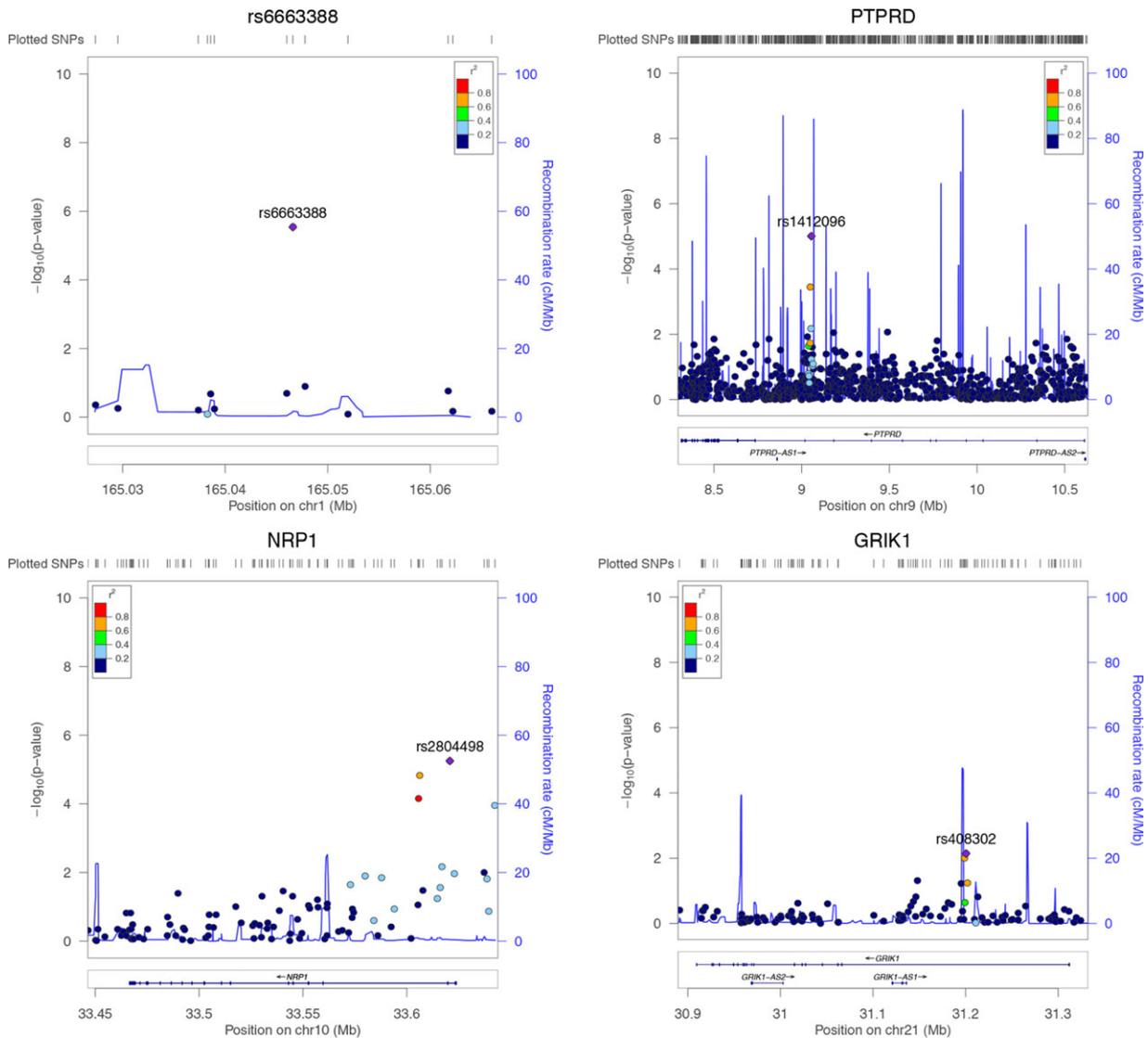
FIGURE 4 LocusZoom for top significant SNPs determined by POM-daSPU for functional connectivity in DMN

3.2.1 | Simulations under varying genetic effect sizes

First, we varied the genetic effect size. The simulation set-up resembled an association pattern between SNP rs2075650 and DMN cortical thicknesses ($p = 12$) in Table 1. We assumed each phenotype to have possibly different inheritance modes: additive or dominant. Subjects were classified into three 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 $\mu_j = (\mu_{j1}, \ldots, \mu_{jp})'$ be a vector for phenotype means for subject group $j \in \{0, 1, 2\}$. Figure 9A illustrates the mean cortical thicknesses of the 12 DMN regions for each genotype group
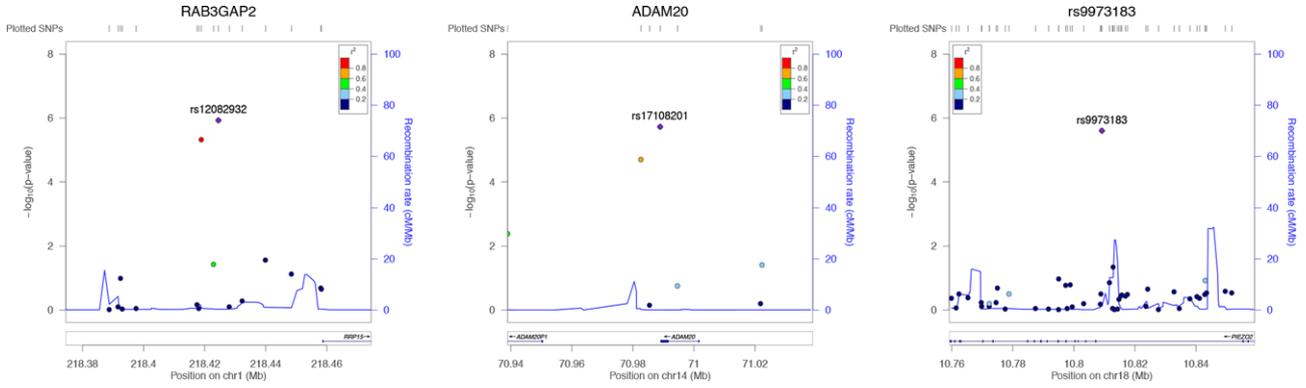
FIGURE 5 LocusZoom for top significant SNPs determined by TATES for functional connectivity in DMN
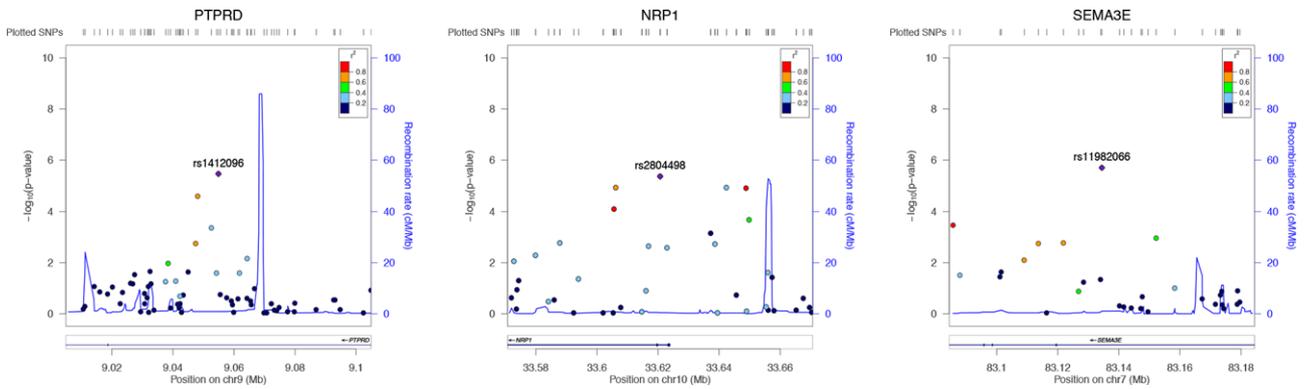
FIGURE 6 LocusZoom for top significant SNPs determined by MDMR for functional connectivity in DMN

TABLE 3 Top significant SNPs for functional connectivity in DMN

rs ID	chr	Nearest gene	position	SPU(1)	SPU(2)	SPU(4)	SPU(∞)	POM-aSPU	$\hat{\gamma}$	POM-daSPU	$(\hat{\gamma}, \hat{t})$	TATES	MDMR
rs6663388	1	NA	165046596	1.10e-01	2.00e-05	2.22e-05	3.02e-07	**6.70e-07**	8	2.90e-06	(2, 0.5)	6.97e-05	2.29e-04
rs7276462	21	GRIK1	31429636	4.94e-01	8.30e-03	1.50e-03	3.00e-04	1.00e-03	7	**9.03e-07**	(3, 0.3)	2.98e-03	5.99e-03
rs11982066	7	SEMA3E	82972395	1.16e-01	1.12e-06	1.34e-05	1.42e-05	9.96e-06	2	5.30e-05	(2, 0.0)	1.06e-03	**2.00e-06**
rs1412096	9	PTPRD	9054913	0.076675	4.00e-06	1.30e-05	1.49e-03	1.70e-05	2	9.92e-06	(2, 0.0)	4.09e-03	**3.50e-06**
rs2804498	10	NRP1	33620707	1.90e-01	2.50e-06	1.79e-04	6.33e-03	1.23e-05	2	5.63e-06	(2, 0.2)	8.01e-02	**4.40e-06**
rs12082932	1	RAB3GAP2	218424514	1.06e-01	1.20e-02	4.00e-03	2.00e-03	5.00e-03	5	1.08e-02	(∞, 0.1)	**1.18e-06**	1.49e-02
rs17108201	14	ADAM20	70988931	4.18e-01	4.00e-04	2.00e-04	7.00e-04	5.00e-04	4	4.00e-03	(4, 0.3)	**1.88e-06**	4.49e-04
rs9973183	18	NA	10809196	6.95e-01	5.80e-02	1.20e-02	2.00e-03	6.00e-03	6	4.80e-02	(∞, 0.0)	**2.52e-06**	6.09e-02

as obtained from the ADNI-1 data. To mimic this pattern, we selected seven traits (i.e., traits 1, 3, 4, 7, 8, 9, 12) to have an additive inheritance mode while the other five traits to have a dominant one. Figure 9B 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 $\beta_j = (\beta_{j1}, \dots, \beta_{jp})'$ 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 $X_i = (X_{i1}, \dots, X_{ip})'$ 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 introduce 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, $\phi = 1$ set the same association strength as that in the real data.
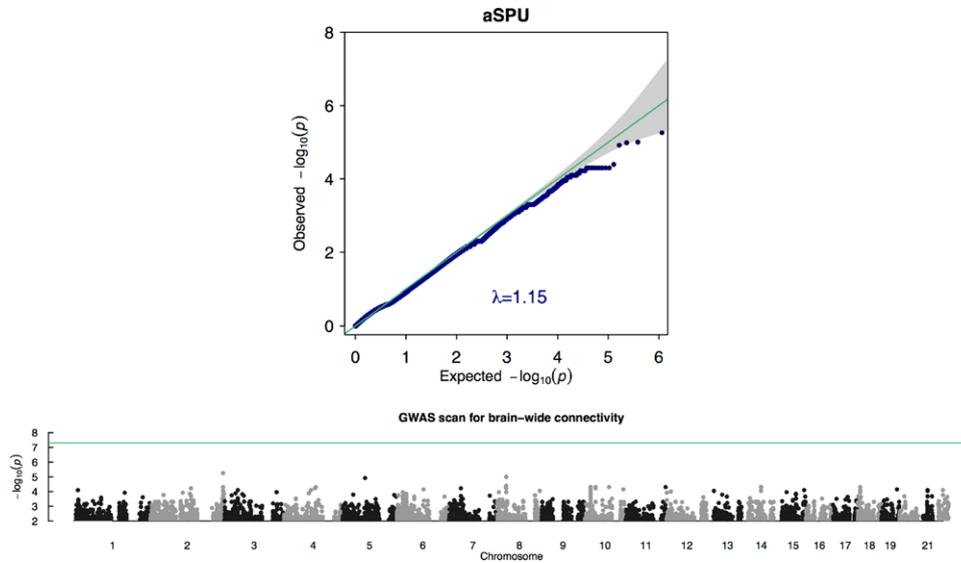
FIGURE 7 QQ plot and Manhattan plot from GWAS for brain-wide function connectivity

TABLE 4 Top significant SNPs from POM-aSPU test for brain-wide functional connectivity

rs ID	chr	Nearest gene	position	SPU(1)	SPU(2)	SPU(4)	SPU(∞)	POM-aSPU	$\hat{\gamma}$
rs11694455	2	NA	239600087	5.83e-02	4.87e-02	1.11e-04	3.02e-07	5.52e-06	∞
rs2937720	5	SV2C	75564181	2.96e-01	2.10e-05	1.8e-05	3.33e-04	1.21e-05	3
rs6981562	8	FAM183CP	29744518	1.61e-07	2.78e-01	2.94e-01	6.33e-03	9.86e-06	1
rs2122068	8	FAM183CP	29745193	8.23e-07	1.35e-01	1.49e-01	7.19e-02	1.03e-05	1
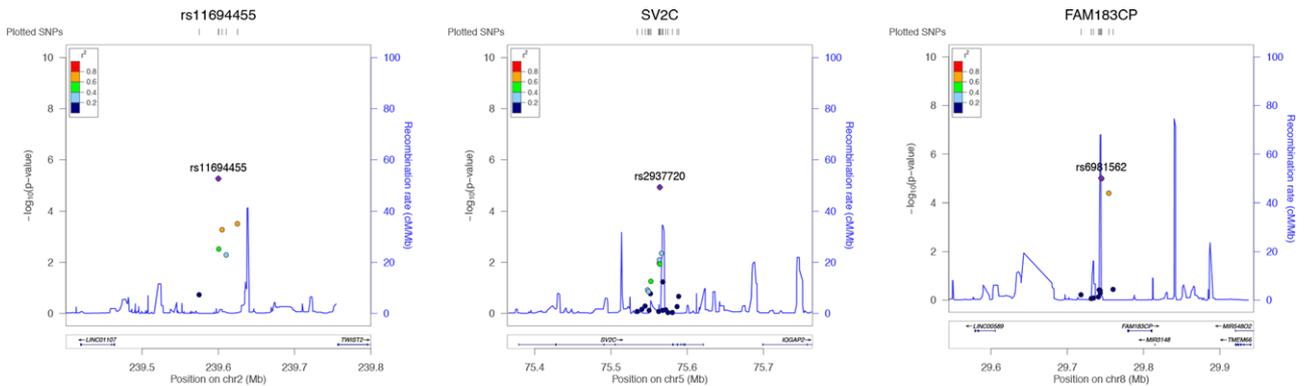
FIGURE 8 LocusZoom for top significant SNPs determined by POM-aSPU for brain-wide functional connectivity

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 2. Among 68 phenotypes, we designated 59 phenotypes to have a dominant inheritance mode, while five to have an additive one, and the remaining five traits were always not associated with the SNP. For set-up 2, MAF was defined at 0.3.

3.2.2 | Simulations under varying proportions of associated phenotypes

In order to evaluate each method's performance in more realistic situations, we varied the proportion of phenotypes associated with the SNP to be tested. A POM was fitted to the original ADNI data with SNP rs429358 and the cortical thicknesses from all brain regions ($p = 68$) to obtain the regres-

sion coefficient estimates. Denote the parameter estimates in the POM, a_j as an intercept for $j = 0, 1$; $b = (b_1, \ldots, b_p)'$ as the regression coefficients in which b_j represents the effect size of the SNP on trait j. First, the multiple phenotypes (X_i) were generated from the multivariate normal distribution with the sample mean and covariance matrix estimated from the ADNI data. The cumulative probability $Pr(Y_i \leq j)$ was calculated based on the inverse logit function: $\exp(a_j + X_i b)/[1 + \exp(a_j + X_i b)]$, from which $\pi_{ij} = Pr(Y_i = j)$ was estimated. The genotype data (0, 1, or 2) was sampled from the multinomial distribution with parameters $\pi_i = (\pi_{i0}, \pi_{i1}, \pi_{i2})$ until a predefined number of each group was reached. For example, out of total 1,000 samples, each group size was predefined in a way that $\sum_{i=1}^{1,000} I(Y_i = 0) = 500$, $\sum_{i=1}^{1,000} I(Y_i = 1) = 380$ and $\sum_{i=1}^{1,000} I(Y_i = 2) = 120$, which ensured the MAF of the
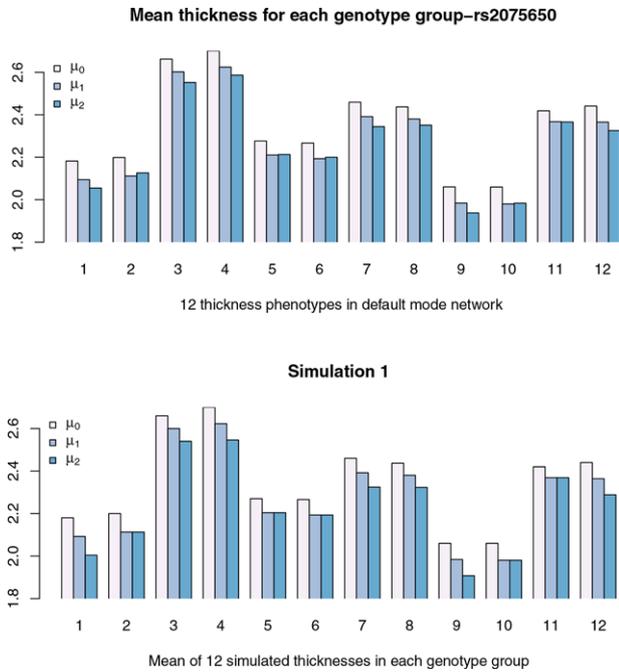
FIGURE 9 Mean phenotype in default mode network and simulation 1

We also conducted a simulation using functional connectivity as a higher dimensional phenotype as encountered in neuroimaging studies. Each effect size b_j was marginally obtained from the POM using SNP rs2804498 and DMN network (described in Real data example). First $p = 30$ of functional connectivity was set causal phenotypes, and multiple null traits were increasingly added to vary the sparsity levels of the association pattern; the phenotypic dimension was only increased up to $p = 140$, because other statistical methods suffered from rank deficiency even with sample size 1,000. As before, the scaling parameter ϕ was employed to yield comparative results. The multiple phenotypes were generated from the multivariate normal distribution and the genotype data was simulated from the multinomial distribution.

To demonstrate performance when $n < p$, we set sample size at 200, and increased the phenotypic dimension up to $p = 1,400$, among which only 100 (or 0) phenotypes were associated with the SNP under the alternative (or null) hypothesis. We applied POM-score, simulation- and permutation-based POM-aSPU, TATES, and MDMR (while other tests were not applicable).

3.2.3 | Type I error and power

Tables 5, 6, 7, 8, and 9 report type I error rates ($\phi = 0$) and power ($\phi > 0$) for each simulation set-up. Type I error rates were well controlled by all methods except MultiPhen (O'Reilly et al., 2012). The inflated type I error rates of

simulated genotype to be around 0.3. We varied the proportion of the associated phenotypes by forcing some of the regression coefficients (b_j) to be zeros; the null phenotypes were randomly selected in each simulation. The effect size of the SNP was controlled with the scaling parameter ϕ.

TABLE 5 Simulation 1: type I errors ($\phi = 0$) and power ($\phi > 0$) under varying genetic effect sizes for 12 phenotypes

	POM			GEE			O'Reilly	van der Sluis		McArdle
ϕ	Score	sim-aSPU	perm-aSPU	Score	sim-aSPU	perm-aSPU	MultiPhen	TATES	MANOVA	MDMR
0	0.053	0.042	0.042	0.057	0.041	0.043	0.059	0.048	0.059	0.050
0.1	0.058	0.056	0.058	0.061	0.065	0.064	0.062	0.068	0.063	0.066
0.2	0.090	0.139	0.140	0.096	0.140	0.138	0.095	0.154	0.096	0.154
0.3	0.164	0.279	0.283	0.168	0.298	0.287	0.168	0.302	0.170	0.321
0.5	0.491	0.690	0.689	0.503	0.703	0.705	0.507	0.732	0.504	0.741
0.7	0.839	0.937	0.935	0.859	0.947	0.947	0.850	0.957	0.860	0.958
1	0.995	0.998	0.999	0.996	1.000	1.000	0.998	0.999	0.997	1.000

sim-aSPU represents the simulation-based aSPU test; perm-aSPU represents the permutation- based aSPU test.

TABLE 6 Simulation 2: type I errors ($\phi = 0$) and power ($\phi > 0$) under varying genetic effect sizes for 68 phenotypes

	POM			GEE			O'Reilly	van der Sluis		McArdle
ϕ	Score	sim-aSPU	perm-aSPU	Score	sim-aSPU	perm-aSPU	MultiPhen	TATES	MANOVA	MDMR
0	0.052	0.046	0.044	0.042	0.054	0.049	0.087	0.041	0.049	0.053
0.1	0.084	0.078	0.077	0.073	0.073	0.084	0.136	0.068	0.084	0.069
0.2	0.239	0.157	0.157	0.205	0.153	0.164	0.336	0.157	0.220	0.136
0.3	0.592	0.332	0.332	0.537	0.311	0.314	0.690	0.334	0.553	0.254
0.4	0.923	0.558	0.569	0.871	0.536	0.533	0.954	0.600	0.883	0.476
0.5	0.997	0.765	0.772	0.994	0.737	0.731	0.999	0.817	0.994	0.680
0.7	1.000	0.966	0.971	1.000	0.959	0.963	1.000	0.988	1.000	0.960

sim-aSPU represents the simulation-based aSPU test; perm-aSPU represents the permutation- based aSPU test.

TABLE 7 Simulation 3: type I errors ($\phi = 0$) and power ($\phi > 0$) under varying association sparsity levels

	#	#	#	POM			GEE			O'Reilly	van der Sluis		McArdle
ϕ	Total	Causals	Nulls	Score	sim-aSPU	perm-aSPU	Score	sim-aSPU	perm-aSPU	MultiPhen	TATES	MANOVA	MDMR
0	68	0	68	0.054	0.049	0.046	0.051	0.045	0.044	0.064	0.038	0.057	0.042
0.4	68	68	0	0.758	0.525	0.528	0.753	0.523	0.518	0.846	0.546	0.770	0.525
		53	15	0.500	0.498	0.500	0.532	0.518	0.510	0.575	0.540	0.546	0.522
		45	23	0.472	0.518	0.526	0.497	0.534	0.538	0.551	0.561	0.515	0.561
		38	30	0.432	0.503	0.509	0.438	0.533	0.541	0.514	0.552	0.448	0.554
		27	41	0.413	0.475	0.482	0.430	0.512	0.514	0.475	0.510	0.448	0.512
		12	56	0.229	0.365	0.363	0.238	0.395	0.393	0.274	0.393	0.242	0.397
0.5	68	68	0	0.975	0.728	0.731	0.972	0.730	0.730	0.990	0.798	0.980	0.751
		53	15	0.702	0.586	0.583	0.727	0.608	0.605	0.752	0.638	0.734	0.605
		45	23	0.681	0.601	0.611	0.703	0.621	0.626	0.726	0.678	0.711	0.641
		38	30	0.615	0.608	0.618	0.645	0.645	0.637	0.675	0.657	0.660	0.650
		27	41	0.540	0.580	0.583	0.588	0.613	0.618	0.633	0.661	0.606	0.631
		12	56	0.329	0.441	0.435	0.339	0.453	0.461	0.392	0.481	0.344	0.476
0.6	68	68	0	0.990	0.800	0.801	0.994	0.821	0.823	0.997	0.888	0.996	0.844
		53	15	0.713	0.645	0.645	0.767	0.676	0.674	0.779	0.709	0.775	0.662
		45	23	0.688	0.622	0.617	0.751	0.645	0.649	0.768	0.725	0.763	0.669
		38	30	0.692	0.638	0.636	0.744	0.674	0.670	0.742	0.706	0.751	0.672
		27	41	0.621	0.643	0.640	0.662	0.671	0.674	0.688	0.698	0.676	0.674
		12	56	0.403	0.501	0.501	0.418	0.531	0.531	0.474	0.550	0.435	0.548

sim-aSPU represents the simulation-based aSPU test; perm-aSPU represents the permutation- based aSPU test.

Multiphen were also discussed in Guo et al. (2015) and Aschard et al. (2014); the inflation became worse with the increasing phenotype dimension (Table 8).

Tables 5 and 6 show the results when all phenotypes were associated with the SNP to be tested. When using $p = 12$ phenotypes (Table 5), the majority of which (7 out of 12) were with the additive inheritance mode, MDMR and TATES performed the best, closely followed by GEE- and POM-based aSPU tests; in general, MultiPhen, MANOVA and GEE- and POM-based score tests performed similarly with power much lower than others in this case. Yet in Table 6 with $p = 68$ phenotypes (most with the dominant inheritance mode), MDMR performed the worst, while MANOVA and POM- and GEE-based score tests showed the highest power; TATES, closely followed by POM- and GEE-based aSPU tests, performed between. Note that the Type I error rate of MultiPhen was inflated with only a moderate number of phenotypes in this case.

Tables 7 and 8 present the results for power under varying association sparsity levels, while Table 9 is for $n < p$. In Table 7, when the number of null phenotypes was small, the score tests, MultiPhen and MANOVA performed similarly and best, followed by TATES, then closely by MDMR and aSPU. However, as the proportion of the associated phenotypes decreased, the power of TATES and MDMR became highest, closely followed by aSPU, all with much higher power than other tests. Table 8 illustrates the similar scenario with a higher dimensional trait: as the proportion of the

associated phenotypes decreased, the relative performance of MDMR, then aSPU, improved consistently and became much more powerful than the other methods. The performance of each test for the $n < p$ set-up is illustrated in Table 9. The power of POM-score test was extremely low (not shown). The general pattern in power was similar to that in Table 9; in particular, MDMR performed best, followed by aSPU. As discussed in Zhang et al. (2014) and Kim et al. (2016), the performance of SPU(2) was close to that of MDMR.

In summary, throughout the simulations, the power of POM-aSPU was very close to that of GEE-aSPU, while POM-score, GEE-score, MANOVA, and MultiPhen performed similarly, though MultiPhen has inflated Type I error rates with a moderate or high number of phenotypes. The comparative performance of a method depended on the simulation setting, i.e., the underlying true but unknown association pattern. Although none of the test was uniformly most powerful, except in some cases when the score test (or similarly MANOVA or MultiPhen) performed best, the aSPU test was either the winner or close to the winner.

3.2.4 | Computing time

Table 10 summarizes the mean computing time for each dataset in simulations in Table 8. The POM-based tests were insensitive to the phenotype dimension, while GEE-based tests required dramatically increasing computing time as the number of phenotypes involved increased. Overall, the

TABLE 8 Simulation 4: type I errors ($\phi = 0$) and power ($\phi > 0$) under varying association sparsity levels for higher dimensional phenotypes

ϕ	# Total	# Causals	# Nulls	POM Score	sim-aSPU	perm-aSPU	GEE Score	sim-aSPU	perm-aSPU	O'Reilly MultiPhen	van der Sluis TATES	MANOVA	McArdle MDMR
0	30	0	30	0.056	0.052	0.054	0.054	0.064	0.066	0.068	0.060	0.056	0.066
	40	0	40	0.040	0.047	0.052	0.037	0.050	0.049	0.050	0.040	0.039	0.054
	50	0	50	0.053	0.050	0.051	0.053	0.053	0.055	0.087	0.060	0.062	0.049
	60	0	60	0.047	0.060	0.055	0.046	0.068	0.064	0.081	0.065	0.050	0.054
	70	0	70	0.044	0.049	0.048	0.041	0.042	0.051	0.087	0.063	0.049	0.051
	80	0	80	0.034	0.044	0.042	0.036	0.046	0.051	0.091	0.046	0.041	0.048
	90	0	90	0.044	0.051	0.052	0.048	0.047	0.051	0.101	0.056	0.057	0.048
	100	0	100	0.034	0.060	0.054	0.034	0.060	0.061	0.113	0.060	0.043	0.052
	110	0	110	0.043	0.046	0.045	0.038	0.050	0.049	0.132	0.054	0.047	0.040
	120	0	120	0.027	0.038	0.043	0.030	0.042	0.049	0.135	0.043	0.041	0.043
	130	0	130	0.040	0.054	0.058	0.037	0.064	0.064	0.150	0.066	0.051	0.057
	140	0	140	0.015	0.038	0.038	0.015	0.042	0.044	-	0.045	-	0.050
0.1	30	30	0	0.920	0.999	0.999	0.909	0.997	0.997	0.932	0.984	0.911	1.000
	40	30	10	0.777	0.983	0.986	0.766	0.987	0.984	0.807	0.925	0.775	0.990
	50	30	20	0.575	0.949	0.950	0.559	0.949	0.947	0.647	0.816	0.576	0.950
	60	30	30	0.429	0.887	0.890	0.421	0.884	0.883	0.516	0.698	0.442	0.893
	70	30	40	0.363	0.841	0.836	0.343	0.835	0.844	0.460	0.606	0.364	0.853
	80	30	50	0.287	0.730	0.741	0.284	0.742	0.750	0.427	0.556	0.312	0.772
	90	30	60	0.270	0.727	0.730	0.257	0.725	0.728	0.438	0.519	0.288	0.746
	100	30	70	0.229	0.681	0.681	0.224	0.661	0.670	0.390	0.493	0.251	0.700
	110	30	80	0.182	0.603	0.612	0.183	0.606	0.615	0.380	0.443	0.202	0.668
	120	30	90	0.179	0.593	0.597	0.177	0.603	0.606	0.391	0.436	0.202	0.642
	130	30	100	0.172	0.573	0.585	0.162	0.585	0.588	0.382	0.440	0.196	0.642
	140	30	110	0.070	0.571	0.588	0.068	0.575	0.593	-	0.445	-	0.644
0.2	60	30	30	0.990	1.000	1.000	0.990	1.000	1.000	0.993	0.999	0.990	1.000
	70	30	40	0.976	0.999	0.999	0.976	0.999	0.999	0.986	0.995	0.978	1.000
	80	30	50	0.941	0.999	0.999	0.945	0.999	0.999	0.974	0.993	0.949	0.999
	90	30	60	0.920	0.999	0.999	0.913	0.999	0.999	0.955	0.993	0.919	0.998
	100	30	70	0.875	0.995	0.996	0.867	0.994	0.994	0.938	0.982	0.883	0.997
	110	30	80	0.814	0.988	0.989	0.806	0.987	0.987	0.915	0.971	0.828	0.993
	120	30	90	0.766	0.987	0.989	0.763	0.987	0.989	0.907	0.961	0.791	0.991
	130	30	100	0.742	0.985	0.986	0.738	0.987	0.986	0.887	0.945	0.761	0.992
	140	30	110	0.585	0.984	0.984	0.574	0.984	0.982	-	0.970	-	0.987
0.3	80	30	50	0.993	1.000	1.000	0.991	1.000	1.000	1.000	0.998	0.993	1.000
	90	30	60	0.988	0.999	0.999	0.984	0.999	1.000	0.995	0.997	0.986	1.000
	100	30	70	0.981	1.000	1.000	0.979	1.000	1.000	0.996	0.999	0.984	1.000
	110	30	80	0.962	0.999	0.999	0.961	0.999	0.999	-	0.998	-	1.000

sim-aSPU represents the simulation-based aSPU test; perm-aSPU represents the permutation- based aSPU test.

proposed POM-score and POM-aSPU tests are much faster than GEE-based tests, and are computationally feasible for high-dimensional phenotypes.

4 | DISCUSSION

We have presented a new adaptive association test for multiple trait-single SNP associations in a proportional odds model. From the analyses of the ADNI data and simulated data, we observed that the POM-aSPU test was competitive as compared to existing tests; in particular, it outperformed many other tests when the proportion of associated phenotypes was low for high-dimensional phenotypes (see Table 9). Neuroimaging phenotypes often are high dimensional in the range of hundreds and more, and in realistic situations, only a subset of the phenotypes are likely to be associated

TABLE 9 Simulation 5: type I errors ($\phi = 0$) and power ($\phi > 0$) under varying association sparsity levels for $n < p$

| ϕ | # Total | # Causals | # Nulls | POM | | | | van der Sluis | McArdle |
				sim-SPU(2)	sim-aSPU	perm-SPU(2)	perm-aSPU	TATES	MDMR
0	300	0	300	0.040	0.044	0.050	0.055	0.063	0.047
	500	0	500	0.054	0.052	0.062	0.054	0.051	0.060
	700	0	700	0.026	0.041	0.040	0.047	0.056	0.039
	900	0	900	0.030	0.047	0.039	0.046	0.063	0.039
	1100	0	1100	0.044	0.054	0.052	0.056	0.058	0.056
	1300	0	1300	0.035	0.041	0.046	0.041	0.070	0.048
	1500	0	1500	0.054	0.053	0.064	0.059	0.056	0.063
0.05	300	100	200	0.457	0.363	0.482	0.394	0.312	0.477
	500	100	400	0.419	0.296	0.437	0.316	0.269	0.431
	700	100	600	0.246	0.180	0.262	0.203	0.176	0.271
	900	100	800	0.326	0.232	0.347	0.252	0.198	0.341
	1100	100	1000	0.266	0.201	0.291	0.232	0.196	0.296
	1300	100	1200	0.216	0.166	0.242	0.191	0.148	0.250
	1500	100	1400	0.228	0.170	0.256	0.179	0.156	0.251
0.10	300	100	200	0.944	0.887	0.946	0.897	0.841	0.951
	500	100	400	0.921	0.849	0.930	0.869	0.791	0.939
	700	100	600	0.750	0.657	0.761	0.687	0.560	0.769
	900	100	800	0.842	0.780	0.859	0.796	0.683	0.862
	1100	100	1000	0.758	0.666	0.782	0.687	0.584	0.766
	1300	100	1200	0.652	0.546	0.690	0.572	0.467	0.694
	1500	100	1400	0.641	0.546	0.681	0.572	0.457	0.676

sim-aSPU represents the simulation-based aSPU test; perm-aSPU represents the permutation-based aSPU test.

TABLE 10 Mean computing time (in seconds) for one dataset in simulations reported in table 8

| ϕ | # Total | # Causals | # Nulls | POM | | | GEE | | | O'Reilly | van der Sluis | | McArdle |
				Score	sim-aSPU	perm-aSPU	Score	sim-aSPU	perm-aSPU	MultiPhen	TATES	MANOVA	MDMR
0.1	30	30	0	0.003	0.093	1.716	3.460	3.514	6.905	0.389	0.033	0.024	43.682
	40	30	10	0.005	0.110	2.230	5.032	5.105	10.260	0.509	0.047	0.027	37.551
	50	30	20	0.006	0.122	2.267	7.992	8.082	13.421	0.614	0.068	0.034	34.772
	60	30	30	0.008	0.136	2.316	12.493	12.601	17.547	0.737	0.096	0.042	32.665
	70	30	40	0.010	0.148	2.378	17.820	17.945	21.848	0.868	0.133	0.049	31.573
	80	30	50	0.012	0.162	2.416	24.267	24.409	26.574	1.008	0.176	0.062	30.029
	90	30	60	0.015	0.176	2.445	32.828	32.989	32.508	1.176	0.230	0.070	29.165
	100	30	70	0.019	0.193	2.505	43.617	43.796	39.753	1.358	0.283	0.082	28.465
	110	30	80	0.022	0.210	2.587	56.435	56.633	47.817	1.567	0.354	0.096	27.971
	120	30	90	0.027	0.229	2.659	71.788	72.004	56.981	1.746	0.448	0.109	27.601
	130	30	100	0.032	0.251	2.743	96.044	96.280	67.910	1.916	0.553	0.128	27.634
	140	30	110	0.038	0.271	2.850	119.949	120.209	81.636	-	0.674	-	28.194

sim-aSPU represents the simulation-based aSPU test; perm-aSPU represents the permutation-based aSPU test.

with an SNP or a gene. Hence the POM-aSPU test can be a useful and powerful method for identifying associations for high dimensional phenotypes. The performance of the POM-aSPU test was similar to that of the GEE-aSPU test (Tables 5, 6, 7, and 8), as supported by an analysis of their similar score vectors (Comparison with existing tests); however, the POM-aSPU test is more robust to the assumed inheritance

mode, and more importantly, is computationally more efficient than GEE-aSPU (Table 10). Moreover, the proposed POM-aSPU test is easily applicable to the high dimensional setting ($n < p$), for which functional connectivity phenotypes were employed as an example (Section Real data example). In the example, we noticed that some, but not all, detected associations likely came from accumulating weak effects of

individual traits (Tables 3 and 4), as the optimal $\hat{\gamma}$ was chosen from SPU(1) or SPU(2) for POM-aSPU test. On the other hand, some associations were detected with $\hat{\gamma} = 8$ or $\hat{\gamma} = \infty$ (Tables 3 and 4), suggesting that POM-aSPU test could also identify one or few traits associated with the SNP with relatively large effect sizes. These results confirmed the adaptiveness of the proposed POM-aSPU test in identifying both joint weak associations and sparse strong signals, both of which may appear but are unknown in practice.

We emphasize that in the current context there is no uniformly most powerful test; the performance of a test depends on the true but unknown association patterns. Hence, although no test can maintain the highest power across all the scenarios, the aSPU test aims to remain powerful across a wide range of situations by adaptively combining over multiple SPU tests. We used the minP or Tippett's method to combine multiple SPU tests, since we expect often only one or few of the SPU tests would be powerful in a given situation. Other combining methods, as discussed in Winkler et al. (2016) may be explored. Furthermore, due to the good performance of the score test (or MANOVA) in some situations, we can combine the POM-based aSPU and score tests as in GEE (Kim et al., 2016; Zhang et al., 2014). Currently, we use permutations or simulations to calculate the *P*-value for the aSPU test, which, albeit feasible, is time-consuming to achieve high significance levels; a parametric approximation to its null distribution along the line of Zhu et al. (2015) would speed it up, but remains to be investigated.

One advantage of the proposed POM-based tests is their straightforward application to a set of mixed types of traits. It would be interesting to apply the proposed POM-aSPU test to such a case. In addition, although we have focused on the study of the proposed POM-aSPU test for association analysis of multiple traits and a single SNP, it can be applied in many other situations. For example, we can apply it to detect association between an ordinal trait (e.g., a disease status like (AD, MCI, CN)) and a group of SNPs. Furthermore, we may extend it to detect associations between multiple traits and multiple SNPs: as in a GEE framework (Kim et al., 2016), we can have a group cumulative logit models, one for each SNP as a response while treating the traits as predictors. These are interesting topics for future investigation.

R package `POMaSPU` implementing the proposed POM-based tests will be posted on CRAN. The GEE-based tests are available in R package `GEEaSPU` on CRAN.

ACKNOWLEDGMENTS

REFERENCES

Aschard, H., Vilhjalmsson, B. J., Greliche, N., Morange, P. E., Tregouet, D. A., & Kraft, P. (2014). Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *The American Journal of Human Genetics*, *94*, 662–676.

Bigos, K. L., Hariri, A. R., & Weinberger, D. R. (2016). Neuroimaging genetics: Principles and practices. New York: Oxford University Press.

Brocker, C. N., Vasiliou, V., & Nebert, D. W. (2009). Evolutionary divergence and functions of the ADAM and ADAMTS gene families. *Human Genomics*, *4*, 43–55.

Damoiseaux, J. S., & Greicius, M. D. (2009). Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Structure and Function*, *213*, 525–533.

Damoiseaux, J. S., Seeley, W. W., Zhou, J., Shirer, W. R., Coppola, G., Karydas, A., … Greicius, M. D.; Alzheimer's Disease Neuroimaging Initiative. (2012). Gender modulates the APOE ϵ4 effect in healthy older adults: convergent evidence from functional brain connectivity and spinal fluid tau levels. *Journal of Neuroscience*, *32*, 8254–8262.

Donix, M., Burggren, A. C., Scharf, M., Marschner, K., Suthana, N. A., Siddarth, P., … Bookheimer, S. Y. (2013). APOE associated hemispheric asymmetry of entorhinal cortical thickness in aging and Alzheimer's disease. *Psychiatry Research*, *214*, 212–20.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., …. Ronald, J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*, 968–980.

Du, L., Huang, H., Yan, J., Kim, S., Risacher, S. L., Inlow, M., … Shen, L., for the Alzheimers Disease Neuroimaging Initiative (2016). Structured sparse

canonical correlation analysis for brain imaging genetics: an improved GraphNet method. *Bioinformatics*, *32*, 1544–1551 doi: 10.1093/bioinformatics/btw033

Ferreira, M. A., & Purcell, S. M. (2009). A multivariate test of association. *Bioinformatics*, *25*, 132–133.

Gutiérrez-Galve, L., Lehmann, M., Hobbs, N. Z., Clarkson, M. J., Ridgway, G. R., Crutch, S., … Barnes, J. (2009). Patterns of cortical thickness according to APOE genotype in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *28*, 476–485.

Glahn, D. C., Winkler, A. M., Kochunov, P., Almasy, L., Duggirala, R., Carless, M. A., … Blangero, J. (2010). Genetic control over the resting brain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 1223–1228.

Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *PNAS*, *101*, 4637–4642.

Guo, X., Li, Y., Ding, X., He, M., Wang, X., & Zhang, H. (2015). Association tests of multiple phenotypes: ATeMP. Chen Z, ed. PLoS ONE 10(10):e0140348. doi:10.1371/journal.pone.0140348

He, Q., Avery, C. L., & Lin, D. Y. (2013). A general framework for association tests with multivariate traits in large-scale genomics studies. *Genetic Epidemiology*, *37*, 759–767.

Hill-Burns, E. M., Singh, N., Ganguly, P., Hamza, T. H., Montimurro, J., Kay, D. M., … Payami, H. (2013). A genetic basis for the variable effect of smoking/nicotine on Parkinson's disease. *Journal of Pharmacogenomics*, *6*, 530–537.

Hirata, Y., Zai, C. C., Souza, R. P., Lieberman, J. A., Meltzer, H. Y., & Kennedy, J. L. (2012). Association study of GRIK1 gene polymorphisms in schizophrenia: case-control and family-based studies. *Human Psychopharmacology*, *27*, 345–351.

Jahanshad, N., Rajagopalan, P., Hua X., Hibar, D. P., Nir, T. M., Toga, A. W., … Thompson, P. M.; the Alzheimer's Disease Neuroimaging Initiative. (2013). Genome-wide scan of healthy human connectome discovers *SPON1* gene variant influencing dementia severity. *PNAS*, *110*, 4768–4773.

Kim, J., Wozniak, J. R., Mueller, B. A., Shen, X., & Pan, W. (2014). Comparison of statistical tests for group differences in brain functional networks. *Neuroimage*, *101*, 681–694.

Kim, J., Wozniak, J. R., Mueller, B. A., & Pan, W. (2015a). Testing group differences in brain functional connectivity: using correlations or partial correlations?. *Brain Connectivity*, *5*, 214–231.

Kim, J., & Pan, W. (2015b). Highly adaptive tests for group differences in brain functional connectivity. *Neuroimage Clinical*, *9*, 625–639.

Kim, J., Zhang, Y., & Pan, W.; for the Alzheimer's Disease Neuroimaging Initiative. (2016). Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data. *Genetics*, *203*, 715–731.

Klei, L., Luca, D., Devlin, B., & Roeder, K. (2008). Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genetics Epidemiology*, *32*, 9–19.

Lesnick, T. G., Papapetropoulos, S., Mash, D. C., Ffrench-Mullen, J., Shehadeh, L., de Andrade, M., … Maraganore, D. M. (2007). A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genetics*, *3*, e98.

Lin, D. Y., & Tang, Z. Z. (2011). A general framework for detecting disease associations with rare variants in sequencing studies. *The American Journal of Human Genetics*, *89*, 354–367.

Lin, J. A., Zhu, H., Knickmeyer, R., Styner, M., Gilmore, J., & Ibrahim, J. G. (2012). Projection regression models for multivariate imaging phenotype. *Genetics Epidemiology*, *36*, 631–641.

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

Liu, Y., Paajanen, T., Westman, E., Wahlund, L. O., Simmons, A., Tunnard, C., … Soininen, H.; AddNeuroMed Consortium. (2010). Effect of APOE ϵ4 allele on cortical thicknesses and volumes: the AddNeuroMed study. *Journal of Alzheimers Diseases*, *21*, 947–966.

McArdle, B. H., & Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, *82*, 290–297.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society Series B*, *42*, 109–142.

Medland, S. E., Jahanshad, N., Neale, B. M., & Thompson, P. M. (2014). Whole-genome analyses of whole-brain data: working within an expanded search space. *Nature Neurosciences*, *17*, 791–800.

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

O'Reilly, P. F., Hoggart, C. J., Pomyen, Y., Calboli, F. C. F., Elliott, P., Jarvelin, M. R., Coin, L. J. (2012). MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE*, *7*, e34861.

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

Passow, S., Specht, K., Adamsen, Tom C., Biermann, M., Brekke, N., Craven, A. R., … Hugdahl, K. (2015). Default-mode network functional connectivity is closely related to metabolic activity. *Human Brain Mapping*, *36*, 2027–2038.

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., … Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, *26*, 2336–2337.

Querbes, O., Aubry, F., Pariente, J., Lotterie, J. A., Démonet, J. F., Duret, V., … Celsis, P.; Alzheimer's Disease Neuroimaging Initiative. (2009). Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain*, *132*, 2036–2047.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, *52*, 1059–1069.

Schifano, E., Li, L., Christiani, D., & Lin, X. (2013). Genome-wide association analysis for multiple continuous secondary phenotypes. *The American Journal of Human Genetics*, *92*, 744–759.

Shen, L., Kim, S., Risachera, S. L., Nho, K., Swaminathan, S., West, J. D.,… Saykin, A. J.; the Alzheimer's Disease Neuroimaging Initiative. (2010). 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*, 1051–1063.

Shen, L., Thompson, P. M., Potkin, S. G., Bertram, L., Farrer, L. A., Foroud, T. M., … Saykin, A. J.; Alzheimers Disease Neuroimaging Initiative. (2014). Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behaviors*, *8*, 183–207.

Shibata, H., Joo, A., Fujii, Y., Tani, A., Makino, C., Hirata, N., … Fukumaki, Y. (2001). Association study of polymorphisms in the GluR5 kainate receptor gene (GRIK1) with schizophrenia. *Psychiatric Genetics*, *11*, 139–144.

Spang, N., Feldmann, A., Huesmann, H., Bekbulat, F., Schmitt, V., Hiebel, C., … Behl, C. (2014). RAB3GAP1 and RAB3GAP2 modulate basal and rapamycin-induced autophagy. *Autophagy*, *10*, 2297–2309.

Sun, Q., Zhu, H., Liu, Y., & Ibrahim, J. G.; for the Alzheimer's Disease Neuroimaging Initiative. (2015). SPReM: sparse projection regression model for high-dimensional linear regression. *Journal of the American Statistical Association*, *110*, 289–302.

Thompson, P. M., Tian, G., Glahn, D. C., Jahanshad, N., & Nichols, T. E. (2013). Genetics of the connectome. *NeuroImage*, *80*, 475–488.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., … Drevets, W. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, *8*, 153–182.

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

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

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

van der Sluis, S., Posthuma, D., & Dolan, C. V. (2013). TATES: efficient multivariate genotype- phenotype analysis for genome-wide association studies. *PLoS Genetics*, 9, e1003235.

van Eimeren, T., Monchi, O., Ballanger, B., & Strafella, A. P. (2009). Dysfunction of the default mode network in Parkinson disease: a functional magnetic resonance imaging study. *Archives of Neurology*, 66, 877–883.

Venkova, K., Christov, A., Kamaluddin, Z., Kobalka, P., Siddiqui, S., & Hensley, K. (2014). Semaphorin 3A signaling through neuropilin-1 is an early trigger for distal axonopathy in the SOD1G93A mouse model of amyotrophic lateral sclerosis. *Journal of Neuropathology and Experimental Neurology*, 73, 702–713.

Vounoua, M., Nichols, T. E., & Montana, G.; the Alzheimer's Disease Neuroimaging Initiative. (2010). Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. *Neuroimage*, 53, 1147–1159.

Walton, E., Geisler, D., Hass, J., Liu, J., Turner, J., Yendiki, A., … Ehrlich, S. (2013). The impact of genome-wide supported schizophrenia risk variants in the neurogranin gene on brain structure and function. *PLoS ONE*, 8, e76815.

Wang, K., & Abbott, D. (2007). A principal components regression approach to multilocus genetic association studies. *Genetic Epidemiology*, 32, 108–118.

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

Wang, Z., Sha, Q., & Zhang, S. (2016). Joint analysis of multiple traits using "optimal" maximum heritability test. *PLoS ONE*, 11, e0150975.

Winkler, A. M., Webster, M. A., Brooks, J. C., Tracey, I., Smith, S. M., & Nichols, T. E. (2016). Non-parametric combination and related permutation tests for neuroimaging. *Human Brain Mapping*, 37, 1486–1511.

Zhang, W., Zhang, Z., Li, X., & Li, Q. (2015). Fitting proportional odds model to case-control data with Incorporating Hardy-Weinberg equilibrium. *Scientific Reports*, 5, 17286.

Zhang, Y., Xu, Z. Shen, X., & Pan, W.; for the Alzheimers Disease Neuroimaging Initiative. (2014). Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*, 96, 309–325.

Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., … Redline, S. (2015). Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *The American Journal of Human Genetics*, 96, 21–36.

APPENDIX A

The Score Vector for POM

Since we were not aware of the availability of a closed-form expression for the score vector for POM, we followed the reparametrization approach of McCullagh (1980) to derive it.

Consider a single multinomial observation $(n_0, \ldots, n_{J-1})$ where $n_j = \sum_{i=1}^{n} I(Y_i = j)$, $j = \{0, \ldots, J-1\}$, and $\sum_{j=0}^{J-1} n_j = n$. Denote $\pi_j = Pr(Y_i = j)$. The likelihood function is $\pi_0^{n_0} \cdots \pi_{J-1}^{n_{J-1}}$. McCullagh (1980) reparametrized the likelihood in terms of a cumulative probability $r_k = \sum_{j=0}^{k} \pi_j$. Define $R_k = \sum_{j=0}^{k} n_j$ and $Z_k = R_k/n$ where $R_{J-1} = n$ and $Z_{J-1} = 1$. The log likelihood can be written as the sum of $J - 1$ quantities

$$l = n \left[\left\{ Z_0 \phi_0 - Z_1 g(\phi_0) \right\} + \left\{ Z_1 \phi_1 - Z_2 g(\phi_1) \right\} \right. \\ \left. + \cdots + \left\{ Z_{J-2} \phi_{J-2} - g(\phi_{J-2}) \right\} \right],$$

where $\phi_j = \text{logit}(r_j/r_{j+1})$ and $g(\phi_j) = \log\{r_{j+1}/(r_{j+1} - r_j)\}$. Recall the cumulative logit model in equation (1):

$$\text{logit}[Pr(Y_i \leq j)] = \text{logit}(r_j) = H_{i(j)}\theta,$$

with $\theta = (\alpha', \delta', \beta')'$ and $H_{i(j)} = (0, \ldots, 1, \ldots, 0, Z_i, X_i)$ where the 1 occurs in the position $j + 1$ where $j \in \{0, \ldots, J-2\}$. 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 follows,

$$\frac{\partial l}{\partial \theta_w} = \sum_{j=0}^{J-2} \frac{\partial l}{\partial \phi_j} \frac{\partial \phi_j}{\partial r_j} \frac{\partial r_j}{\partial \theta_w},$$

$$\frac{\partial l}{\partial \phi_j} = R_j - R_{j+1} \frac{r_j}{r_{j+1}},$$

$$\frac{\partial \phi_j}{\partial r_j} = r_{j+1}/\{r_j(r_{j+1} - r_j)\},$$

$$\frac{\partial r_j}{\partial \theta_w} = r_j(1 - r_j)H_{i(j,w)} - r_j(1 - r_{j+1})H_{i(j+1,w)}.$$

Considering individual level probabilities, the score vector is defined by

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

with $r_{ij} = Pr(Y_i \leq j)$ and $\phi_{ij} = \text{logit}\{r_{ij}/r_{i(j+1)}\}$. Each component of $U_\theta = (U'_\alpha, U'_\delta, U'_\beta)'$ arrives

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

$$U_\delta = \sum_{i=1}^{n} \sum_{j=0}^{J-2} (1 - r_{i(j-1)} - r_{ij}) \cdot I(Y_i = j) \cdot Z_i,$$

$$U_{\alpha g} = \sum_{i=1}^{n} \frac{1 - r_{ig}}{r_{i(g+1)} - r_{ig}} \left[I(Y_i \le g) r_{i(g+1)} - I\{Y_i \le (g+1)\} r_{ig} \right]$$

$$- \frac{1 - r_{ig}}{r_{ig} - r_{i(g-1)}} \left[I\{Y_i \le (g-1)\} r_{ig} - I(Y_i \le g) r_{i(g-1)} \right],$$

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

APPENDIX B

The Covariance Matrix of the Score Vector

The covariance matrix of the score vector, $Cov(U_\theta) = [A_{ws}]$, can be obtained as

$$A_{ws} = \sum_{i=1}^{n} \sum_{j=0}^{J-2} \frac{r_{i(j+1)}}{r_{ij}(r_{i(j+1)} - r_{ij})} q_{jw}(i) q_{js}(i),$$

$$q_{jw}(i) = r_{ij}(1 - r_{ij}) H_{i(j,w)} - r_{ij}(1 - r_{i(j+1)}) H_{i(j+1,w)}.$$