



Published in final edited form as:

Neuroimage. 2017 February 01; 146: 983–1002. doi:10.1016/j.neuroimage.2016.09.055.

Genome-wide Association Analysis of Secondary Imaging Phenotypes from the Alzheimer's Disease Neuroimaging Initiative Study

Wensheng Zhu^{a,b}, Ying Yuan^d, Jingwen Zhang^b, Fan Zhou^b, Rebecca C Knickmeyer^c, and Hongtu Zhu^{b,e,*} for the Alzheimer's Disease Neuroimaging Initiative[†]

^aSchool of Mathematics & Statistics and KLAS Northeast Normal University, Changchun 130024, China

^bDepartment of Biostatistics, University of North Carolina at Chapel Hill Chapel Hill, NC 27599, USA

^cDepartment of Psychiatry, University of North Carolina at Chapel Hill Chapel Hill, NC 27599, USA

^dTakeda Pharmaceuticals U.S.A., Inc. 300 Massachusetts ave, Cambridge, MA 02139

^eDepartment of Biostatistics University of Texas MD Anderson Cancer Center Houston, TX 77030, USA

Abstract

The aim of this paper is to systematically evaluate a biased sampling issue associated with genome-wide association analysis (GWAS) of imaging phenotypes for most imaging genetic studies, including the Alzheimer's Disease Neuroimaging Initiative (ADNI). Specifically, the original sampling scheme of these imaging genetic studies is primarily the retrospective case-control design, whereas most existing statistical analyses of these studies ignore such sampling scheme by directly correlating imaging phenotypes (called the secondary traits) with genotype. Although it has been well documented in genetic epidemiology that ignoring the case-control sampling scheme can produce highly biased estimates, and subsequently lead to misleading results and suspicious associations, such findings are not well documented in imaging genetics. We use extensive simulations and a large-scale imaging genetic data analysis of the Alzheimer's Disease Neuroimaging Initiative (ADNI) data to evaluate the effects of the case-control sampling scheme on GWAS results based on some standard statistical methods, such as linear regression methods, while comparing it with several advanced statistical methods that appropriately adjust for the case-control sampling scheme.

*Address for correspondence and reprints: Hongtu Zhu, Ph.D., hzhu5@mdanderson.org; Phone No: 346-814-0191.

[†]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

Keywords

Alzheimer's Disease Neuroimaging Initiative; Case-control; Genome-wide association analysis; Imaging genetics; Secondary trait

1 Introduction

The case-control design has been widely used in many imaging studies, such as Alzheimer's Disease Neuroimaging Initiative (ADNI). The case-control study usually starts with identifying two or more groups with the known outcome (e.g., case and control) and then retrospectively identifies risk factors (e.g., structural and functional imaging) that may contribute to the outcome. Case-control studies are comparatively quick, inexpensive, and easy, but they provide less evidence for causal inference than a randomized controlled trial. They are particularly appropriate for investigating outbreaks and studying rare diseases or outcomes. For example, although the overall design of ADNI is a longitudinal prospective study for studying various biomarkers at baseline and their longitudinal profiles, ADNI is essentially a case-control design for studying genetic influences on these biomarkers, since the ADNI participants is not a random sample of the age-matched general population (Kim and Pan, 2015).

The use of the case-control design in imaging genetic studies raises many intricate statistical issues in the joint analysis of imaging, genetic, clinical, and cognitive data (Lin and Zeng, 2009; Monsees et al., 2009; Schifano et al., 2013; Wei et al., 2013; Chen et al., 2013; Ghosh et al., 2013; Tchetgen, 2014; Kim and Pan, 2015). Specifically, in case-control imaging genetic studies, there are four sets of variables including marker genotype(s), G , secondary traits, Y , primary (outcome) phenotype, D , and clinical variables X . Imaging measures have been widely used as secondary (or intermediate) traits, that may be directly associated with a specific disease outcome, for most neuropsychiatric and neurodegenerative illnesses (Chou et al., 2009; Filippini et al., 2009; Jahanshad et al., 2010; Yoon et al., 2010; Molina et al., 2011; Kremen et al., 2010; Montag et al., 2008; Chiang et al., 2011; Gilmore et al., 2010; Shen et al., 2010; Peterson et al., 2009). The key difficulty comes from the fact that both secondary traits Y and marker genotype(s) G are collected conditional on the primary phenotype D , whereas the main target of inference is the population model for Y given G . As a result, it may be essential to adjust for D when one models Y given G in case-control studies.

It is well known in genetic epidemiology that for quantitative trait, improperly handling case-control sampling scheme can lead to estimation bias, inflated false positive rate, and decreased power (Lin and Zeng, 2009; Tchetgen, 2014; Kim and Pan, 2015), while one may gain substantial powers by appropriately accounting for the case-control scheme (Lin and Zeng, 2009; Monsees et al., 2009; Schifano et al., 2013; Wei et al., 2013; Chen et al., 2013; Ghosh et al., 2013; Tchetgen, 2014). To the best of our knowledge, however, existing GWAS of imaging phenotypes either ignore the case-control sampling scheme or include both diagnosis and the interaction between diagnosis and SNP as covariates (Potkin et al., 2009). Until recently, Kim and Pan (2015) wrote a cautionary note on the potential importance of

adjusting the case-control design in the GWAS of imaging measures. Specifically, they compared the retrospective likelihood method in Lin and Zeng (2009) and the standard linear regression method by using simulation studies and ADNI data analysis and concluded that standard linear regression is generally valid (with only small biases or slightly inflated Type I errors) for the ADNI data.

The aim of this paper is to understand the effects of the case-control sampling scheme on GWAS of imaging measures. There are two major contributions in this paper. The first one is to carry out GWAS of imaging measures obtained from ADNI by comparing statistical methods with and without correcting for the case-control study design (Tchetgen, 2014; Lin and Zeng, 2009). To this end, we first study the association between 501,584 Single Nucleotide Polymorphisms (SNPs) and each of 93 regions of interest (ROIs) across 362 subjects including 198 cognitively normals (CNs) and 164 AD patients from ADNI-1. Then to confirm the results with a larger sample size, we combine ADNI-1 with ADNI-2 in order to study the association between 6,017,259 markers (including SNPs and indels) and the right hippocampus across 494 subjects (CN and AD patients). In contrast, in Kim and Pan (2015), they focused on the genetic markers on the 19th chromosome and single ROI volume (right hippocampus), while they did not adjust for population stratification. The second one is to carry out additional simulation studies based on relatively simple genetic models in order to examine the case-control sampling scheme. We consider some settings, in which moderate association exists between the secondary trait and disease status. In contrast, in Kim and Pan (2015), they focused on the settings, in which a weak correlation exists between the secondary trait and disease status. Specifically, in their simulations, the odds ratio of the secondary trait with disease status is very close to one. Therefore, it is not surprising that simple linear regression only introduces small amount of biases and slightly increase inflated Type I errors.

2 Methods

Suppose that imaging genetic data are collected from n independent subjects under the retrospective case-control sampling design. For each subject, given the case-control status D_i (0 or 1), we observe the imaging measure Y_i of interest, the clinical factors X_i as well as the genotype score G_i at one of SNP markers along the whole genome for $i = 1, \dots, n$. We only consider the additive mode of inheritance, where G_i counts the number of copies of the minor allele at the locus. We assume, from now on, that Y_i is a quantitative trait due to the fact that almost all imaging measures of interest are continuous. In the following, we briefly describe two sets of methods to model the association between single genetic variant and imaging phenotype with and without considering the case-control sampling scheme, both of which can be used to screen the whole genome in GWAS of imaging phenotypes.

2.1 Methods without Handling Case-control Sampling Scheme

Most imaging genetic studies to date do not account for the case-control sampling scheme in the association analysis of imaging phenotype Y_i and genetic variant G_i , but rather include the case-control status D_i as a covariate or exclude it in the association model.

Linear regression model without adjusting for disease status—The most popular method is to regress Y_i on G_i with only X_i as covariates. Specifically, we consider a linear regression model as follows:

$$Y_i = \alpha_0 + \alpha_1 G_i + \alpha_2 X_i + \varepsilon_i, \quad (1)$$

where α_0 , α_1 , and α_2 are regression coefficients, and ε_i 's are measurement errors with zero mean and variance σ^2 .

Linear regression model adjusting for disease status—The other method is to regress Y_i on G_i while including both X_i and D_i as covariates in the model. Specifically, we consider a linear regression model as follows:

$$Y_i = \alpha_0 + \alpha_1 G_i + \alpha_2 X_i + \alpha_3 D_i + \varepsilon_i, \quad (2)$$

where α_3 is a regression coefficient. We may also consider an interaction model by adding the interaction between G_i and D_i in the above model (Potkin et al., 2009) as follows,

$$Y_i = \alpha_0 + \alpha_1 G_i + \alpha_2 X_i + \alpha_3 D_i + \alpha_4 G_i \times D_i + \varepsilon_i. \quad (3)$$

In both models (1) and (2), we are interested in making statistical inference on α_1 by using a test statistic, such as score test, to test whether $\alpha_1 = 0$ holds or not. In model (3), we are interested in making statistical inference on α_4 as we did for α_1 in both models (1) and (2). Moreover, α_2 can be a vector due to the presence of multiple environmental factors, among others.

2.2 Methods with Properly Handling Case-control Sampling Scheme

Several approaches have been proposed to properly account for the case-control scheme in the association analysis of secondary trait. Here we adopt two of them for comparison purposes.

A retrospective likelihood method—Lin and Zeng (2009) proposed a retrospective likelihood function $\prod_{i=1}^n P(Y_i, G_i, X_i | D_i)$ given by

$$\prod_{i=1}^n \left\{ \frac{P(D_i=1|X_i, G_i, Y_i)P(Y_i|X_i, G_i)P(X_i, G_i)}{P(D_i=1)} \right\}^{D_i} \times \left\{ \frac{P(D_i=0|X_i, G_i, Y_i)P(Y_i|X_i, G_i)P(X_i, G_i)}{P(D_i=0)} \right\}^{1-D_i},$$

where $P(D_i=1) = \int_Y \int_{G, X} P(D_i=1|X_i, G_i, Y_i)P(Y_i|X_i, G_i)P(X_i, G_i)dYdGdX$, and $P(D_i=0) = 1 - P(D_i=1)$. We model $P(Y_i|X_i, G_i)$ as a linear regression model given by

$$Y_i = \alpha_0 + \alpha_1 G_i + \alpha_2 X_i + \varepsilon_i,$$

while we consider a logistic regression model as follows:

$$P(D_i=1|X_i, G_i, Y_i) = \frac{e^{\gamma_0 + \gamma_1 G_i + \gamma_2 X_i + \gamma_3 Y_i}}{1 + e^{\gamma_0 + \gamma_1 G_i + \gamma_2 X_i + \gamma_3 Y_i}}.$$

We are interested in making inference on α_1 . Since we are not interested in $P(X_i, G_i)$, it is possible to treat $P(X_i, G_i)$'s as nuisance parameters. To make inference on α_1 , Lin and Zeng (2009) proposed a profile-likelihood approach to eliminate the nuisance parameters and obtain a profile-likelihood function, which can be maximized by the Newton-Raphson algorithm. More details please refer to Lin and Zeng (2009). In this paper, we use the software SPREG downloaded from <http://dlin.web.unc.edu/software/spreg-2/>.

A re-parameterization of the conditional model—Tchetgen (2014) presented a general regression framework for the analysis of secondary trait in case-control studies. This method is based on a careful non-parametric reformulation of the conditional model for the secondary trait given D and X . As pointed out by Tchetgen (2014), the inverse probability weighted regression method (IPW) (Monsees et al. (2009)) is a special case of his method and therefore, we do not consider IPW here.

Following Tchetgen (2014), we are mainly interested in two mean models as follows. One is the population mean model for Y_i given (G_i, X_i) given by

$$\mu(G_i, X_i; \boldsymbol{\alpha}) = E(Y_i | G_i, X_i) = \alpha_0 + \alpha_1 G_i + \alpha_2 X_i.$$

The other one is the conditional mean of Y_i given (G_i, X_i, D_i) given by

$$\tilde{\mu}(G_i, X_i, D_i) = E(Y_i | G_i, X_i, D_i) = E(Y_i | G_i, X_i, D_i, S_i=1), \quad (4)$$

where $S_i=1$ indicates selection into the case-control sample and the second equality holds under the assumption that selecting individuals into the case-control study is independent of (Y_i, G_i, X_i) given D_i by using the unmatched case-control design.

The key idea of Tchetgen's (2014) method is to re-parameterize the conditional mean model (4) based on a relationship between $\mu(G_i, X_i; \boldsymbol{\alpha})$ and $\tilde{\mu}(G_i, X_i, D_i)$ as follows:

$$\tilde{\mu}(G_i, X_i, D_i) = \mu(G_i, X_i; \boldsymbol{\alpha}) + \{D_i - p(G_i, X_i)\} H(G_i, X_i), \quad (5)$$

where $p(G_i, X_i) \equiv P(D_i=1 | G_i, X_i)$. Moreover, $H(G_i, X_i) \equiv E(Y_i | G_i, X_i, D_i=1) - E(Y_i | G_i, X_i, D_i=0)$ is often modeled as

$$H(G_i, X_i; \boldsymbol{\beta}) = \beta_0 + \beta_1 G_i + \beta_2 X_i.$$

Furthermore, we have

$$\text{logit}(p(G_i, X_i)) = \text{logit}(P(D_i=1|G_i, X_i, S_i=1)) + \log \frac{\bar{p}(1-\bar{\pi})}{\bar{\pi}(1-\bar{p})},$$

where $\bar{p} = P(D_i=1)$ and $\bar{\pi} = P(D_i=1|S_i=1)$ are the disease prevalences in the target population and in the case-control sample, respectively. Moreover, $\text{logit}(P(D_i=1|G_i, X_i, S_i=1))$ can be modeled as a logistic regression as follows:

$$\text{logit}(P(D_i=1|G_i, X_i, S_i=1; \boldsymbol{\eta})) = \eta_0 + \eta_1 G_i + \eta_2 X_i. \quad (6)$$

The estimation of $\boldsymbol{\eta}$ in model (6) is followed by the estimations of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in model (5). Here, we are also interested in making inference on α_1 . See Tchetgen (2014) for more details.

3 Simulation Studies

We use Monte Carlo simulations to evaluate the finite sample performance of the five methods mentioned in Section 2 including (i) LReg: linear regression method without adjusting for case-control status; (ii) LRegD: linear regression method adjusted for case-control status; (iii) LRegDG: linear regression method adjusted for case-control status as well as interaction between case-control status and genetic factor; (iv) SPREG: the retrospective likelihood method in Lin and Zeng (2009); and (v) SEE: the approach based on the re-parameterization of conditional model in Tchetgen (2014).

3.1 Simulation Setup

The data were simulated through the following steps:

1. Generate design vector $X = (X_1, X_2)^T$, where X_1 is simulated from a Bernoulli distribution $\text{Bernoulli}(0.5)$ and X_2 is simulated from a normal distribution $N(0, 1)$.
2. Generate biallelic genetic variable G . Given a minor allele frequency (MAF) of p_A and assuming Hardy-Weinberg equilibrium, SNP genotypes (AA, Aa and aa) were simulated from a multinomial distribution with frequency $(p_A^2, 2p_A(1-p_A), (1-p_A)^2)$ for (AA, Aa, aa). We consider only the additive mode of inheritance, under which the genetic variable was coded as the number of minor alleles. We set $p_A = 0.1, 0.3, \text{ and } 0.45$.
3. Generate quantitative trait Y from a linear regression model $\alpha_0 + \alpha_1 G + \alpha_2 X + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$. We set $\alpha_0 = \sigma^2 = 1$ and $\alpha_2 = (-0.1, -0.2)$. Then, we set α_1

= 0 under the null hypothesis and $\alpha_1 = -0.12$ under the alternative hypothesis, respectively.

4. Generate case-control status. The case-control status was simulated using a logistic model given by

$$\text{logit}P(D=1|G, Y) = \gamma_0 + \gamma_1 G + \gamma_2 Y.$$

We considered two values of γ_2 : $-\log(2)$ and $\log(2)$ and three values of γ_1 : 0, $\log(1.2)$ and $\log(1.4)$. Then, we chose γ_0 to get a disease prevalence of 10%. Negative γ_2 indicates that the secondary trait and the quantitative trait are negatively correlated. For example, reduced brain volumes is often seen in Alzheimer's Disease patients and reduced gray matter (GM) volume is often seen in subjects with schizophrenia.

5. Repeat Steps 1–4 until a sample with 100,000 is obtained.
6. Sample 1000 cases and 1000 controls from the above large pool of subjects.

Since we focus on the association study of imaging measures, we will primarily calculate Type I error rates and powers for each method instead of assessing estimation biases. For each combination of simulation parameters, we simulated 100,000 data sets and then estimated Type I error rates and powers of the four estimation methods. The estimated Type I error rates and powers are defined as the ratios of the number of tests that were rejected among the 100,000 replications under the null hypothesis ($\alpha_1 = 0$) and the alternative hypothesis ($\alpha_1 = -0.12$), respectively.

3.2 Type I Error Rates

Table 1 shows Type I error rates of the five methods for the analysis of the SNP-secondary trait association at different MAFs and two given nominal significance levels 0.01 and 0.05. If the Type I error rate for a test exceeds the nominal significance level too much, it cannot control Type I error well and is undesirable. The closer that the Type I error rate is to the nominal significance level, the better the test will be. When no SNP is associated with the disease, except for LReg, all other four methods have comparable Type I error rates and can appropriately control Type I errors at the nominal significance level. However, the Type I error rates for LReg increase with the odds ratio of SNP with the case-control status, especially when the secondary trait and the disease status are negatively correlated ($OR_{DY} = 0.5$). For LRegD, although the Type I error rates can be controlled well when the secondary trait and the disease status are negatively correlated, they increase with the odds ratio of SNP when the secondary trait and the disease status are positively correlated ($OR_{DY} = 2$). In contrast, LRegDG, SEE in Tchetgen (2014), and SPREG in Lin and Zeng (2009) maintain appropriate Type I errors in all scenarios.

3.3 Power Results

As shown in Figures 1–2, when no SNP is associated with the disease, LReg, LRegD, and SPREG have similar power, whereas they have higher power than SEE. The power of LRegDG is extremely small for all the simulation settings. Similar to the results for Type I

errors, for all MAF settings, LReg and LRegD are not stable with respect to the odds ratio of SNP and the case-control status. When the secondary trait and the disease status are positively correlated ($OR_{DY}=2$), the power of LReg decreases with the odds ratio of SNP and the case-control status. Similar phenomenon can be found for LRegD when the secondary trait and the disease status are negatively correlated ($OR_{DY}=0.5$). Lin and Zeng (2009) also addressed the power loss in their main conclusions about the standard statistical methods for the secondary trait analysis. Please refer to Appendix B of Lin and Zeng (2009) for more details. For some settings, the higher power of LReg and LRegD comes with the cost of inflated type I error rates. In contrast, the power of SEE and SPREG remains stable regardless the direction of the correlation between the secondary trait and the case-control status, and no matter what the MAF is. This leads to a noteworthy observation that it is very important to appropriately handle the case-control sampling scheme in the analysis of secondary trait.

4 GWAS for ADNI

4.1 ADNI

“Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators 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.”

4.2 Sample

Two data sets, including imaging and genetic data as well as demographics, have been used throughout the paper. We focus on the first data set, which contains 818 ADNI-1 subjects, out of which 806 also had baseline neuroimaging data. For the second data set, we also include 432 ADNI-2 subjects and merge them with ADNI-1 subjects in order to increase the

sample size. Both ADNI data sets contain three groups of subjects classified according to disease status, which are AD patients, patients with mild cognitive impairment (MCI) and cognitively normal (CN) subjects. Since we focus on evaluating how the case-control sampling scheme affects GWAS of ADNI imaging measures, we only consider AD patients and CN subjects from the baseline diagnostic groups for both data sets.

4.3 MRI Acquisition and Image Preprocessing

The MRI data in ADNI 1, collected across a variety of 1.5 Tesla MRI scanners with protocols individualized for each scanner, included standard T1-weighted images obtained using volumetric three-dimensional sagittal MPRAGE or equivalent protocols with varying resolutions. The typical MRI protocol for ADNI1 included the following parameters: repetition time (TR) = 2400 ms, inversion time (TI) = 1000 ms, flip angle = 8°, field of view (FOV) = 24 cm with a $256 \times 256 \times 170$ acquisition matrix in the x-, y-, and z-dimensions yielding a voxel size of $1.25 \times 1.26 \times 1.2$ mm³ (Jack et al. (2008)). All original uncorrected image files are available to the general scientific community at <http://adni.loni.usc.edu/>. All participants newly enrolled in ADNI 2 are scanned using the 3T MRI scanning protocol. The typical MRI protocol for ADNI 2 included the following parameters: 8-channel coil, TR = 400 ms, TE = min full, flip-angle = 11°, slice thickness = 1.2 mm, resolution = 256×256 mm and FOV = 26 cm.

We processed the MRI data by using standard steps as follows. First, AC (anterior commissure) and PC (posterior commissure) correction were performed on all original images using MIPAV software, which is a free image analysis program developed by the National Institutes of Health and is available at <http://mipav.cit.nih.gov> (McAuliffe et al. (2002)). We re-sampled the images to have dimension $256 \times 256 \times 256$. N2 bias field correction was implemented to reduce intensity inhomogeneity in these reconstructed images (Sled et al. (2002)). For each subject, we aligned the follow-up time point images to the baseline image using rigid registration for one subject in order to keep intracranial registration consistent. Skull-stripping on the baseline image were performed using a hybrid of two widely-used algorithms: Brain Surface Extractor (BSE) (Shattuck et al. (2001)) and Brain Extraction Tool (BET) (Smith (2002)), which can be used to compensate for problems encountered with individual methods and ensure the accuracy of the skull-stripping results. Intensity inhomogeneity correction followed the skull-stripping procedure. Then, we removed the cerebellum from the images based on registration using a manually-labeled cerebellum as a template. We performed intensity inhomogeneity correction for the third time and subsequently segmented the brain into four different tissues: grey matter (GM), white matter (WM), ventricle (VN), and cerebrospinal fluid (CSF) using the FSL-FAST software (Zhang et al. (2001)).

We used the deformation field that we obtained during registration to generate RAVENS maps (Davatzikos et al. (2001); Davatzikos and Resnick (1998); Goldszal et al. (1998)) to quantify the local volumetric group differences for the whole brain and each of the segmented tissue type (GM, WM, VN, and CSF), respectively. Regional volumetric measurements and analyses are then performed via measurements and analyses of the resulting tissue density maps. This technique has previously been applied to a variety of

longitudinal aging studies (Beresford et al. (2006); Fan et al. (2008); Resnick et al. (2000)) and has been extensively validated. Lastly, we carried out automatic regional labeling: first, by labeling the template and second, by transferring the labels following the deformable registration of subject images. Labeling of the ROIs for each subject's data was done automatically and based on previously-validated segmented atlas of the human brain (Kabani et al. (1998)) After labeling 93 ROIs, we were able to compute volumes for each of these ROIs for each subject.

4.4 Genotype data

For the first data set, the subjects from ADNI-1 were genotyped using the Human610-Quad BeadChip (Illumina, Inc. San Diego, CA). The original data contained 620,901 markers, including multiple types of genetic variants. This paper focuses on single nucleotide polymorphisms (SNPs) and only SNP markers on the autosomes were analyzed (582,539 out of 598,821 SNP markers on the autosomes). To reduce the population stratification effect, we only used 749 Caucasians from all 818 subjects with complete imaging measurements at baseline. The following quality control (QC) procedures were then performed: (i) call rate check per subject, (ii) gender check, (iii) sibling pair identification, and (iv) population stratification. After the quality control procedures, 708 subjects remained. Furthermore, SNPs were excluded from the imaging genetics analysis if they could not meet any of the following criteria: (i) call rate per SNP $\geq 95\%$, (ii) minor allele frequency (MAF) $\geq 5\%$, and (iii) Hardy-Weinberg equilibrium test of $p \geq 10^{-6}$. Remaining missing genotype variables were imputed as the modal value. Finally, by removing all MCI subjects, 362 subjects and 501,584 SNPs remained in the final data analysis.

For the second data set, to increase the sample size, we merged ADNI-1 and ADNI-2 together. The subjects from ADNI-2 were genotyped using the Illumina Human OmniExpress BeadChip (Illumina, Inc. San Diego, CA). The original data contained 730,525 markers, including multiple types of genetic variants. We applied the following preprocessing technique to the genetic data obtained from ADNI-1 and ADNI-2. The similar quality control procedures, except that the call rate per SNP was changed to $\geq 90\%$, were performed on the genotype data for ADNI-1 and ADNI-2, separately. After the quality control procedures, 503,892 SNPs obtained from 22 chromosomes of 708 Caucasians with complete imaging measurements at baseline were included in ADNI-1, and 517,152 SNPs of 349 Caucasians with complete imaging measurements at baseline were included in ADNI-2. MACH-Admix software (<http://www.unc.edu/~yunmli/MaCH-Admix/>) (Liu et al., 2013) was applied to perform genotype imputation, using 1000G Phase I Integrated Release Version 3 haplotypes (<http://www.1000genomes.org>) (1000 Genomes Project Consortium, 2012) as a reference panel. The 7,986,566 bi-allelic markers (including SNPs and indels) were included in ADNI-1 and 8,218,182 markers were included in ADNI-2. Finally, the two data sets were merged based on the intersection of markers and quality control was also conducted after imputation, excluding markers with (i) low imputation accuracy (based on imputation output R^2), (ii) Hardy-Weinberg equilibrium $p < 10^{-6}$, and (iii) MAF $< 5\%$. Finally, 494 AD and CN subjects and 6,017,259 bi-allelic markers (including SNPs and indels) remained in the final data analysis.

4.5 GWAS analysis

Since LRegDG has extremely small power according to our simulation studies, we only applied the four methods: LReg, LRegD, SEE, and SPREG mentioned in Section 2 to the 93 ROIs for the first data set, and to the right hippocampus only for the second data set. To correct for population stratification, we adopt the widely used principal component analysis (PCA) approach that was proposed by Price et al. (2006). This approach applies principal components analysis to genotype data in order to identify several top principal components (PCs), which are the continuous axes of genetic variations, and then these PCs are treated as covariates in the association analysis. Here we only consider additive genetic model and include age, sex, and the top 5 principal component scores as non-genetic covariates. In addition, ICV is generally considered to be an accurate indicator of brain volume. In order to adjust for the effect of brain volume, the volume phenotypes were transformed by first being divided by ICV and then taking logarithm as response variables. For SEE (Tchetgen (2014)) and SPREG (Lin and Zeng (2009)), an estimated disease prevalence is needed. However, since the disease prevalence of ADNI varies over time as well as different subpopulation of age based on the annual reports of ADNI, it is not straightforward to estimate the disease prevalence properly for the ADNI case-control sample. We varied the disease prevalence rate from 0.1 to 0.4 with an increment of 0.05.

4.6 Results

For the sake of space, we first present GWAS analysis based on 6 selected ROIs including the left and right amygdala, the left and right temporal lobes, and the left and right hippocampi at different disease prevalences for the first data set. Since the aim of this paper is to evaluate the effects of the case-control sampling scheme on different methods of GWAS, we did not use the usual significance threshold of GWAS due to only very few SNPs passing 5×10^{-8} for all the methods. In order to achieve better comparison results, we use 10^{-6} as the cut point of the p -values although we may also use other cut points for our GWAS of imaging measures. The quantile-quantile (Q-Q) plots of GWAS analyses on the right hippocampus, the left amygdala, and the right amygdala are shown in Figures 3–5. The Q-Q plots for the left hippocampus and the left (or right) temporal lobe are not included here since there are no SNPs with p -value $< 10^{-6}$. Figures 3–5 reveal that SEE is severely affected by disease prevalence (dp), whereas SPREG performs very well under different disease prevalence rates. The Manhattan plots of GWAS analyses on the right hippocampus, the left amygdala, and the right amygdala are shown in Figures 6–8. Here we only include the results corresponding to LReg, LRegD, and SEE for the right hippocampus and the left amygdala as $dp = 0.25$ and for the right amygdala as $dp = 0.2$, and those corresponding to SPREG as $dp = 0.15$ and $dp = 0.35$. All Q-Q plots and Manhattan plots show that LReg and SPREG give similar results for the 6 selected ROIs in the first data set, which are basically consistent with the results of Kim and Pan (2015). Moreover, including disease status as a covariate in the linear model (LRegD) may be problematic for the GWAS analyses of secondary imaging phenotypes from ADNI.

Table 2 presents seven SNP-ROI pairs with their p -values smaller than 10^{-6} . The results corresponding to p -values smaller than 10^{-6} are highlighted if the corresponding method has a reasonable Q-Q plot as shown in Figures 3–5. Moreover, although SEE detects many SNP-

ROI pairs with their p -values smaller than 10^{-6} , we did not include them in Table 2 due to the poor performance of the Q-Q plots. As shown in Table 2, SPREG is able to detect more SNP-ROI pairs with values smaller than 10^{-6} compared with LReg and/or LRegD. Specifically, six SNP-ROI pairs with their p -values smaller than 10^{-6} are identified by SPREG, whereas only four of them are found by LReg, and one is found by the LRegD. SPREG leads to the smallest p -values compared with all other methods, if disease prevalence can be chosen properly. All these results indicate that it is helpful to properly consider the sampling scheme in GWAS of secondary imaging phenotypes even though LReg is seemed to be reasonable. The SNP marker rs2075650, identified by LReg and SPREG, is associated with the right hippocampus and the right amygdala. The association between rs2075650 and the right hippocampus was also detected by previous works (Shen et al. (2010); Xu et al. (2014)).

Figures 9–11 present the LocusZoom plots around SNPs in 19q13 for the right hippocampus and right amygdala, in 18q21.31 and 5q33.1 for the left amygdala, and in 14q24.3 for the right amygdala. In the 18q21.31 region, rs12457258 and rs12605132 are highly correlated with each other ($r^2 > 0.8$) and show similar low p -values for the association with the left amygdala volume, even though rs12605132 does not show significant association at the significance level $< 10^{-6}$. In the 5q33.1 region, rs10476743 and rs1056993 are both highly correlated with rs1135093 ($r^2 > 0.8$), rs4629585 is correlated with rs1135093 ($r^2 > 0.6$), and they show similar low p -values for the association with the left amygdala volume. In the 14q24.3 region, rs2655997 is correlated with rs207801 ($r^2 > 0.6$) and the p -value for the association with the right amygdala volume is $7.32E-6$. In the 19q13 region, rs2075650 has the smallest p -value for the right hippocampus volume and the right amygdala volume.

We summarize some key findings for other 87 ROIs in the first dataset. For SEE and SPREG, we only include the results corresponding to $dp = 0.25$ here. Table 3 lists the numbers of SNP-ROI pairs with their p -values smaller than 10^{-6} that are identified by SPREG and/or SEE, but not by LReg and LRegD for each of other 87 ROIs. Most of these SNP-ROI pairs were identified by SPREG. Table 4 lists the numbers of SNP-ROI pairs with their p -values smaller than 10^{-6} identified by the methods with or without accounting for the sampling scheme, in which the p -values corresponding to the SNP-ROI pairs identified by SPREG and/or SEE are smaller than those identified by LReg and LRegD. If we set 10^{-8} as the significance threshold, Figures 12–15 reveal that SPREG detects more significant SNP-ROI pairs with their p -values smaller than 10^{-8} than LReg and LRegD for the four ROIs, including me.f-o.gy.R, sup.f.gy.R, inf.f.gy.L, and fornix.R. Although LReg also detects two significant SNP-ROI pairs with their p -values smaller than 10^{-8} (Figures 13 and 15), SPREG has smaller p -values than LReg for these two SNP-ROI pairs. These summaries show that SPREG is better than LReg in GWAS of these four ROIs from ADNI-1.

Finally, we present the GWAS results of the right hippocampus in the second data set. Here, we only include the results with $dp = 0.25$ for SEE and the results with $dp = 0.15$ and $dp = 0.35$ for SPREG. Figure 16 shows the Q-Q plots and Manhattan plots of GWAS on the right hippocampus. Again, LReg and SPREG, even SEE, give the similar results.

Overall, LReg is a good choice for most of ROIs from ADNI, but all the above results lead us to believe that it may be necessary to consider the sampling scheme in GWAS of secondary imaging phenotypes from ADNI.

5 Discussion

The two aims of this paper are to draw an attention on the importance of accounting for the case-control sampling scheme in the genome-wide association study of secondary imaging phenotypes and to understand how much the inflated false positive rate may be produced by standard LReg and LRegD when neglecting the sampling scheme and how much power we may potentially gain by properly accounting for it.

We have used the ADNI data to show that LReg and SPREG give very similar results for most of ROIs although LReg does not take account of the sampling scheme, which is probably due to the high prevalence of the AD in the target population as pointed out by Kim and Pan (2015). However, further analyses showed that more significant SNP-ROI pairs with their p -values smaller than 10^{-8} can be identified by using SPREG based on the retrospective likelihood method for me.f.o.gy.R, sup.f.gy.R, inf.f.gy.L and fornix.R. We have used extensive simulations to show that linear regression methods including LReg and LRegD show severely inflated Type I error rates when the genetic variant is associated with the case-control status and reduce powers when the secondary trait is highly associated with the case-control status. Moreover, the retrospective likelihood method performs very well both for power calculation and for controlling of Type I error.

Although we have designed the simulation studies by assuming a single quantitative secondary trait and the additive mode of inheritance, the studies can be easily extended to consider other kind of secondary traits such as binary traits (Wang and Shete (2011); Chen et al. (2013)), longitudinal traits (Skup et al. (2012); Xu et al. (2014)), multiple traits (Lin et al. (2012); Zhang et al. (2014); Zhu et al. (2014)) as well as other modes of inheritance. We only include AD patients and CN subjects in the GWAS of ADNI data in the paper, but we may include the MCI subjects in our GWAS and treat them as controls by following the analysis in Kim and Pan (2015). Moreover, it may be more interesting to develop new methods for secondary imaging phenotypes under the multiple-group design.

Although our conclusions are only based on the analysis of ANDI dataset, in which the main conclusion is that it is helpful to adjust for the sampling scheme in GWAS of secondary imaging phenotypes, we expect that the general conclusions will not be changed too much even for other imaging genetic datasets. Following the discussions of Kim and Pan (2015), it is necessary to properly adjust the sampling scheme of any imaging genetic study when the disease prevalence in the target population is much less than that of AD in ADNI study. The topic initiated by this paper is rather important and timely, as there is an increasing interest in GWAS of secondary imaging data, but almost all existing analyses do not account for the case-control sampling scheme. We hope that our paper along with Kim and Pan (2015) can encourage more scientists to conduct further research on the development of new methods for secondary-trait analysis and other related issues in imaging genetic studies.

Acknowledgments

This material was based upon work partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. The research of the first author was partially supported by NSFC-11371083 and CSC-201406625026. The research of Dr. Zhu was partially supported by NSF grants SES-1357666 and DMS-1407655, NIH grant MH086633, and a grant from Cancer Prevention Research Institute of Texas. The research of Dr. Knickmeyer was supported by NIH grants 1R01MH092335. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH and NSF.

References

- 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. [PubMed: 23128226]
- Beresford T, Arciniegas D, Alfors J, Clapp L, Martin B, Du Y, Liu D, Shen D, Davatzikos C. Hippocampus volume loss due to chronic heavy drinking. *alcoholism. Clinical and Experimental Research*. 2006; 30:1866–1870.
- Chen H, Kittles R, Zhang W. Bias correction to secondary trait analysis with casecontrol design. *Statistics in Medicine*. 2013; 32:1494–1508. [PubMed: 22987618]
- Chiang MC, McMahon KL, de Zubicaray GI, Martin NG, Hickie I, Toga AW, Wright MJ, Thompson PM. Genetics of white matter development: A dti study of 705 twins and their siblings aged 12 to 29. *NeuroImage*. 2011; 54:2308–2317. [PubMed: 20950689]
- Chou Y, Lepore N, Chiang MC, Avedissian C, Barysheva M, McMahon KL, de Zubicaray GI, Meredith M, Wright MJ, Toga AW, Thompson PM. Mapping genetic influences on ventricular structure in twins. *NeuroImage*. 2009; 44:1312–1323. [PubMed: 19041405]
- Davatzikos C, Genc A, Xu D, Resnick S. Voxel-based morphometry using the ravens maps: methods and validation using simulated longitudinal atrophy. *NeuroImage*. 2001; 14:1361–1369.
- Davatzikos C, Resnick S. Sex differences in anatomic measures of interhemispheric connectivity: correlations with cognition in women but not men. *Cerebral Cortex*. 1998; 8:635–640. [PubMed: 9823484]
- Fan Y, Gur R, Gur R, Wu X, Shen D, Calkins M, Davatzikos C. Unaffected family members and schizophrenia patients share brain structure patterns: a high-dimensional pattern classification study. *Biological Psychiatry*. 2008; 63:118–124. [PubMed: 17555720]
- Filippini N, Rao A, Wetten S, Gibson RA, Borrie M, Guzman D, Kertesz A, Loy-English I, Williams J, Nichols T, Whitcher B, Matthews PM. Anatomically-distinct genetic associations of apoe 4 allele load with regional cortical atrophy in alzheimer's disease. *NeuroImage*. 2009; 44:724–728. [PubMed: 19013250]
- Ghosh A, Wright F, Zou F. Unified analysis of secondary traits in case-control association studies. *Journal of the American Statistical Association*. 2013; 108:566–576.
- Gilmore JH, Schmitt JE, Knickmeyer RA, Smith JK, Lin W, Styner M, Gerig G, Neale MC. Genetic and environmental contributions to neonatal brain structure: a twin study. *Human Brain Mapping*. 2010; 31:1174–1182. [PubMed: 20063301]
- Goldszal A, Davatzikos C, Pham D, Yan M, et al. An image-processing system for qualitative and quantitative volumetric analysis of brain images. *Journal of Computer Assisted Tomography*. 1998; 22:827–837. [PubMed: 9754125]
- Jack C, Bernstein M, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*. 2008; 27:685–691. [PubMed: 18302232]
- Jahanshad N, Lee AD, Barysheva M, McMahon KL, de Zubicaray GI, Martin G, Wright J, Toga W, Thompson P. Genetic influences on brain asymmetry: A dti study of 374 twins and siblings. *NeuroImage*. 2010; 52:455–469. [PubMed: 20430102]
- Kabani N, MacDonald D, Holmes C, Evans A. A 3d atlas of the human brain. *Neuroimage*. 1998; 7:S717.
- Kim J, Pan W. A cautionary note on using secondary phenotypes in neuroimaging genetic studies. *NeuroImage*. 2015; 121:136–145. [PubMed: 26220747]
- Kremen W, Prom-Wormley E, Panizzon MS, Eyster LT, Fischl B, Neale MC, Franz CE, Lyons MJ, Pacheco J, Perry MAS, Schmitt JE, Grant MD, Seidman LJ, Thermenos HW, Tsuang MT, Eisen

- SA, Dale AM, Fennema-Notestine C. Genetic and environmental influences on the size of specific brain regions in midlife: The vetsa mri study. *NeuroImage*. 2010; 49:1213–1223. [PubMed: 19786105]
- Lin D, Zeng D. Proper analysis of secondary phenotype data in case-control association studies. *Genetic Epidemiology*. 2009; 33:256–265. [PubMed: 19051285]
- Lin J, Zhu H, Knickmeyer R, Styner M, Gilmore J, Ibrahim J. Projection regression models for multivariate imaging phenotype. *Genetic Epidemiology*. 2012; 36:631–641. [PubMed: 22807230]
- Liu E, Li M, Wang W, Li Y. Mach-admix: Genotype imputation for admixed populations. *Genetic Epidemiology*. 2013; 37(1):25–37. [PubMed: 23074066]
- McAuliffe, M., Lalonde, F., McGarry, D., Gandler, W., Csaky, K., Trus, B. Medical image processing, analysis and visualization in clinical research. in: *Computer-based medical systems. 14th IEEE Symposium on. IEEE; 2002. p. 381-386.*
- Molina V, Papiol S, Sanz J, Rosa A, Arias B, Fatjó-Vilas M, Calama J, Hernández AI, Bècker J, Faòans L. Convergent evidence of the contribution of tp53 genetic variation (pro72arg) to metabolic activity and white matter volume in the frontal lobe in schizophrenia patients. *NeuroImage*. 2011; 56:45–51. [PubMed: 21296169]
- Monsees G, Tamimi R, Kraft P. Genome-wide association scans for secondary traits using case-control samples. *Genetic Epidemiology*. 2009; 33:717–728. [PubMed: 19365863]
- Montag C, Reuter M, Newport B, Elger C, Weber B. The bdnf val66met polymorphism affects amygdala activity in response to emotional stimuli: Evidence from a genetic imaging study. *NeuroImage*. 2008; 42:1554–1559. [PubMed: 18603005]
- Peterson B, Warner V, Bansal R, Zhu H, Hao X, Liu J, Durkin K, Adams P, Wickramaratne P, Weissman M. Cortical thinning in persons at increased familial risk for major depression. *Proc Natl Acad Sci U S A*. 2009; 106:6273–6278. [PubMed: 19329490]
- Potkin S, Guffanti G, Lakatos A, Turner J, Kruggel F, Fallon J, Saykin A, Orro A, Lupoli S, Salvi E, Weiner M, Macciardi F. for the Alzheimer's Disease Neu-roimaging Initiative. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for alzheimer's disease. *PLoS ONE*. 2009; 4(8):e6501. [PubMed: 19668339]
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Re-ich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006; 38(8):904–909. [PubMed: 16862161]
- Resnick S, Goldszal A, Davatzikos C, Golski S, Kraut M, Metter E, Bryan R, Zonderman A. One-year age changes in mri brain volumes in older adults. *Cerebral Cortex*. 2000; 10:464–472. [PubMed: 10847596]
- Schifano E, Li L, Christiani D, Lin X. Genome-wide association analysis for multiple continuous secondary phenotypes. *The American Journal of Human Genetics*. 2013; 92:744–759. [PubMed: 23643383]
- Shattuck D, Sandor-Leahy S, Schaper K, Rottenberg D, Leahy R. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*. 2001; 13:856–876. [PubMed: 11304082]
- Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, Foroud T, Pankratz N, Moore JH, Sloan CD, Huentelman MJ, Craig DW, DeChairo BM, Potkin SG, CRJ, Weiner MW, Saykin AJ. ADNI. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in mci and ad: A study of the adni cohort. *NeuroImage*. 2010; 53:1051–1063. [PubMed: 20100581]
- Skup M, Zhu H, Zhang H. Multiscale adaptive marginal analysis of longitudinal neuroimaging data with time-varying covariates. *Biometrics*. 2012; 68:1083–1092. [PubMed: 22551084]
- Sled J, Zijdenbos A, Evans A. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *Medical Imaging, IEEE Transactions*. 2002; 17:87–97.
- Smith S. Fast robust automated brain extraction. *Human Brain Mapping*. 2002; 17:143–155. [PubMed: 12391568]
- Tchetgen E. A general regression framework for a secondary outcome in case-control studies. *Biostatistics*. 2014; 5:117–128.
- Wang J, Shete S. Estimation of odds ratios of genetic variants for the secondary phenotypes associated with primary disease. *Genetic Epidemiology*. 2011; 35:190–200. [PubMed: 21308766]

- Wei J, Carroll R, Muller U, Keilegom I. Robust estimation for homoscedastic regression in the secondary analysis of case-control data. *Journal of the Royal Statistical Society, Series B.* 2013; 75:185–206.
- Xu Z, Shen X, Pan W. for the Alzheimer’s Disease Neuroimaging Initiative. Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. *PLoS ONE.* 2014; 9(8):e102312. [PubMed: 25098835]
- Yoon U, Fahim C, Perusse D, Evans AC. Lateralized genetic and environmental influences on human brain morphology of 8-year-old twins. *NeuroImage.* 2010; 53:1117–1125. [PubMed: 20074649]
- Zhang Y, Brady M, Smith S. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imagin.* 2001; 20:45–57.
- Zhang Y, Xu Z, Shen X, Pan W. for the ADNI. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage.* 2014; 96:309–325. [PubMed: 24704269]
- Zhu H, Khondker Z, Lu Z, Ibrahim J. Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. *Journal of the American Statistical Association.* 2014; 109:977–990.

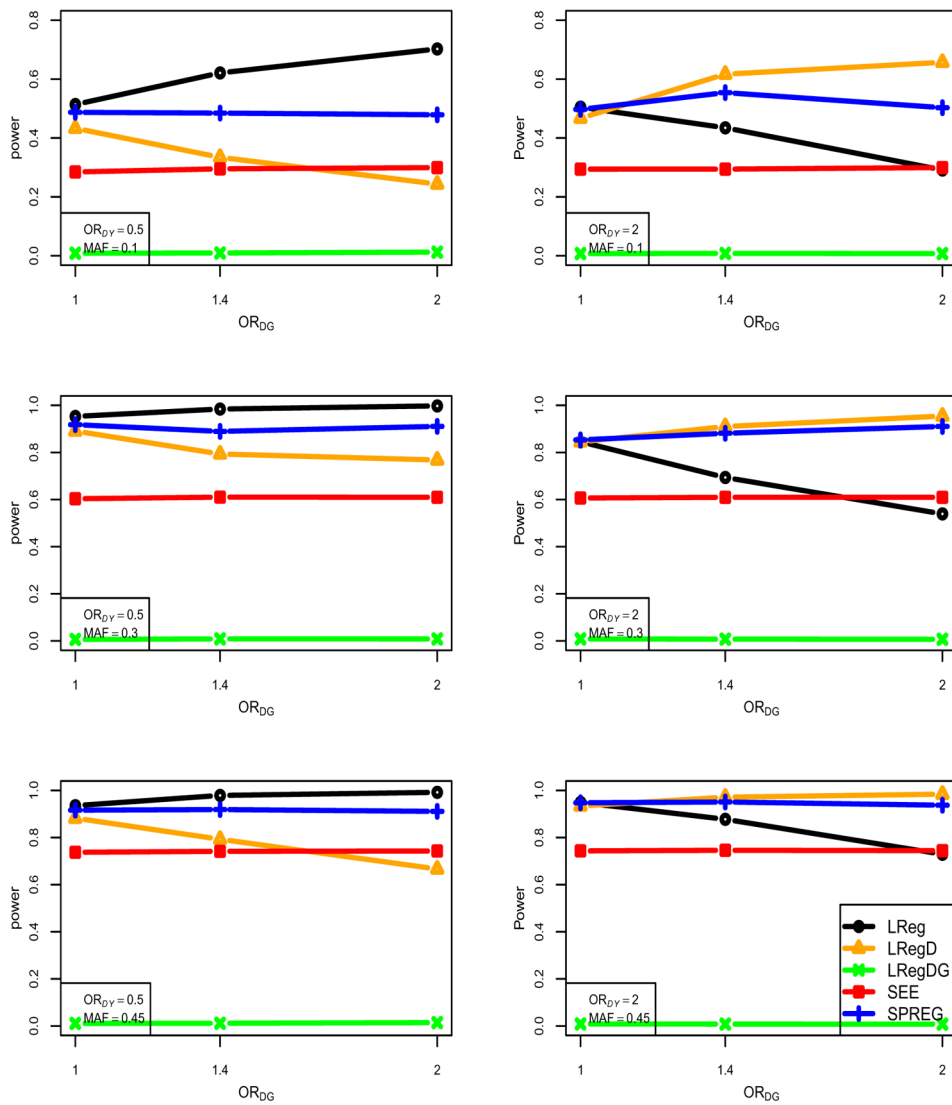


Figure 1. Simulation results: power of association tests at the 0.01 nominal significance level and three MAFs 0.1 (first row), 0.3 (second row), and 0.45 (third row) for the five analysis methods: LReg, LRegD, LRegDG, SEE, and SPREG based on 100,000 simulated data sets for the 10% disease prevalence. OR_{DG} and OR_{DY} refer to the odds ratio of disease with the SNP and the odds ratio of disease with the secondary trait, respectively.

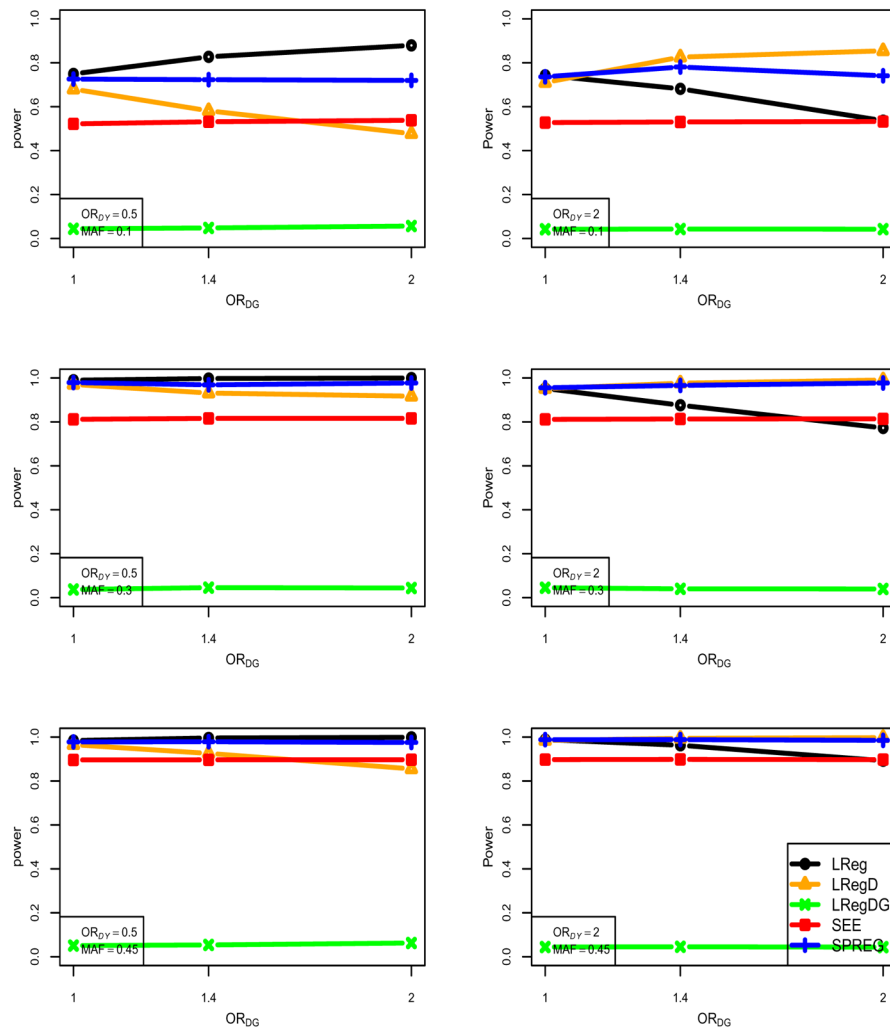


Figure 2. Simulation results: power of association tests at the 0.05 nominal significance level and three MAFs 0.1 (first row), 0.3 (second row), and 0.45 (third row) for the five analysis methods: LReg, LRegD, LRegDG, SEE, and SPREG based on 100,000 simulated data sets for the 10% disease prevalence. OR_{DG} and OR_{DY} refer to the odds ratio of disease with the SNP and the odds ratio of disease with the secondary trait, respectively.

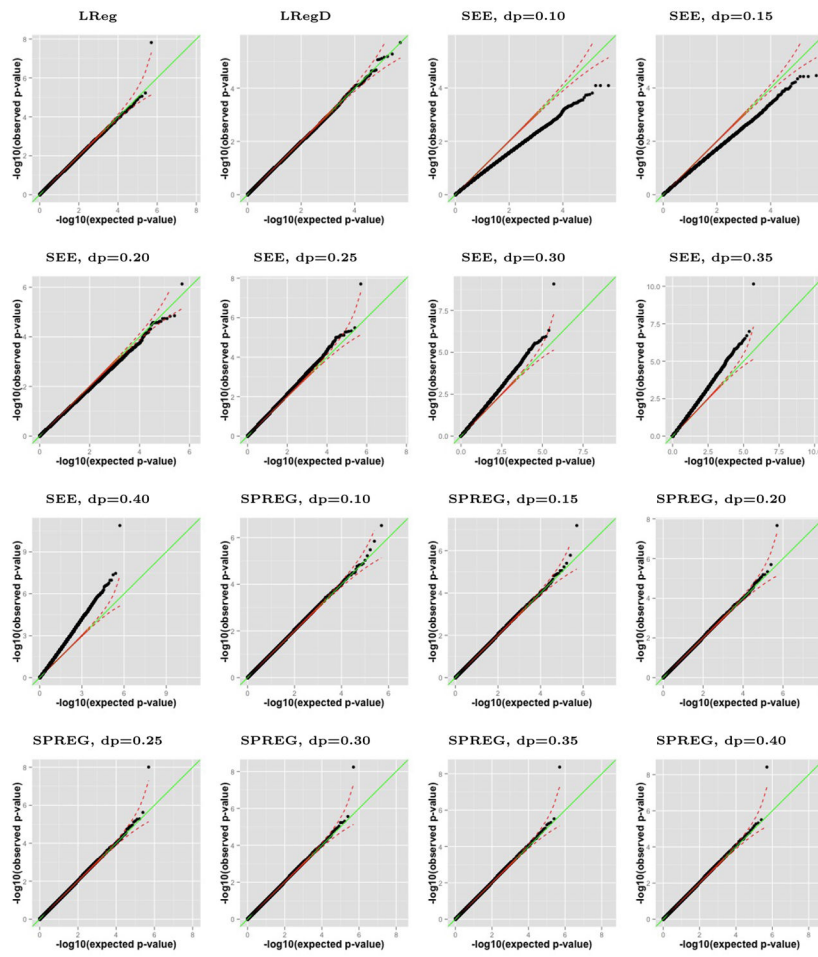


Figure 3. ADNI data analysis results: Q-Q plots of genome-wide association study (GWAS) of the right hippocampus volume by using the four methods including LReg, LRegD, SEE, and SPREG.

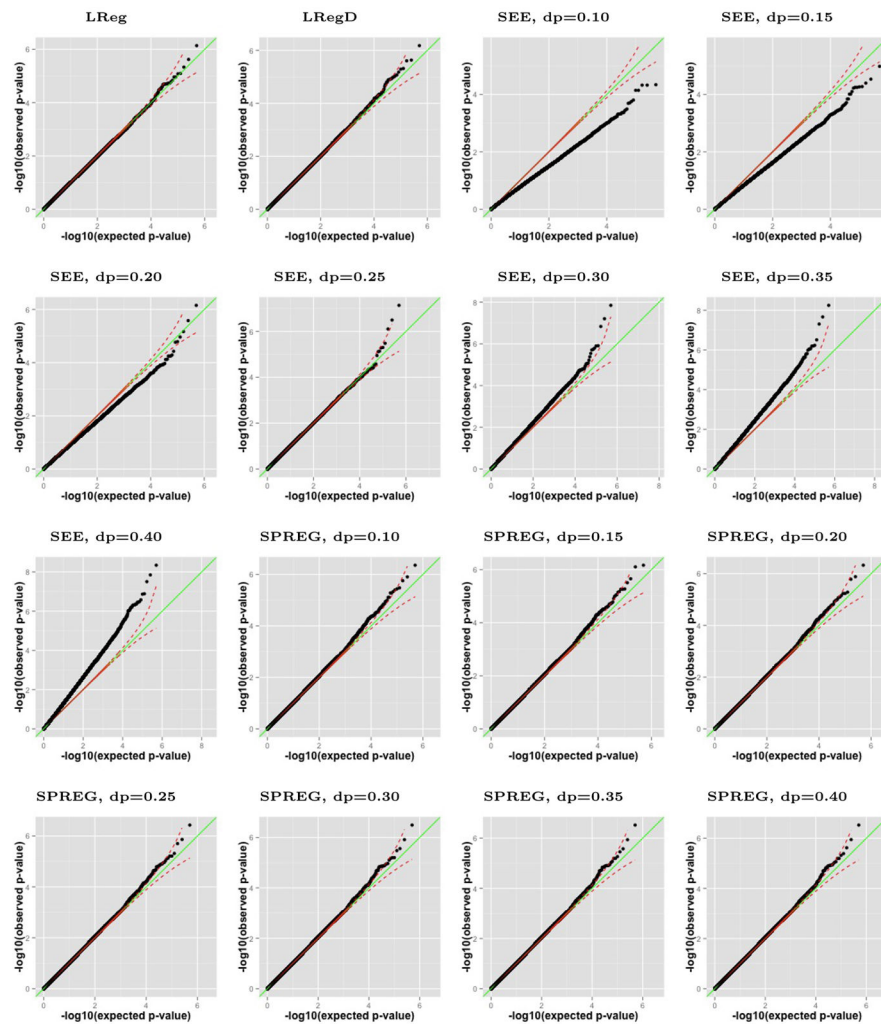


Figure 4. ADNI data analysis results: Q-Q plots of genome-wide association study (GWAS) of the left amygdala volume by using the four methods including LReg, LRegD, SEE, and SPREG.

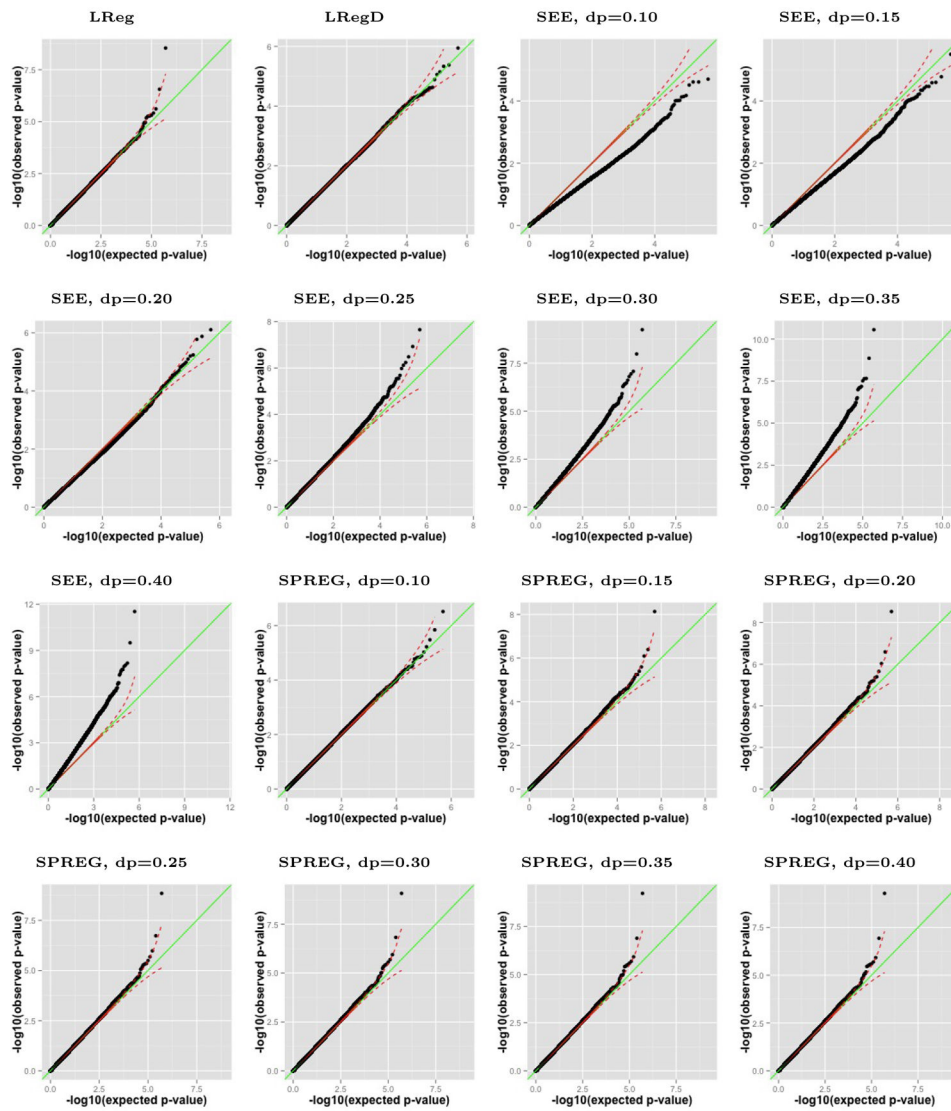


Figure 5. ADNI data analysis results: Q-Q plots of genome-wide association study (GWAS) of the right amygdala volume by using the four methods including LReg, LRegD, SEE, and SPREG.

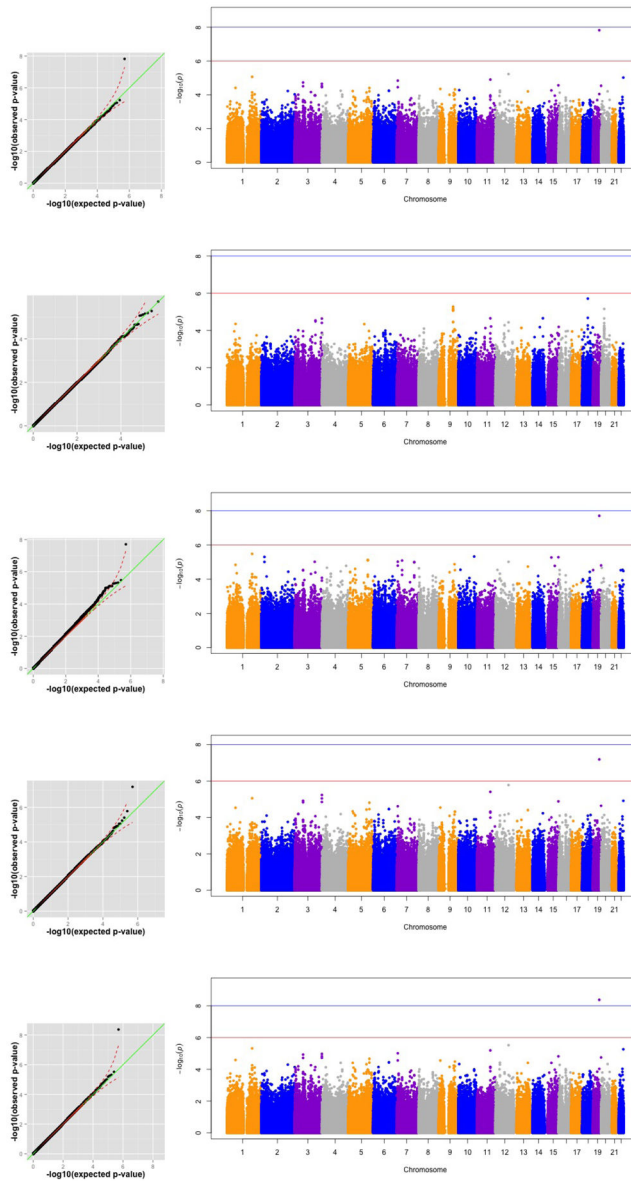


Figure 6. ADNI data analysis results: Q-Q and Manhattan plots of genome-wide association study (GWAS) of the right hippocampus volume by using LReg (first row), LRegD (second row), SEE with $dp = 0.25$ (third row), and SPREG with $dp = 0.15$ (fourth row) and $dp = 0.35$ (fifth row).

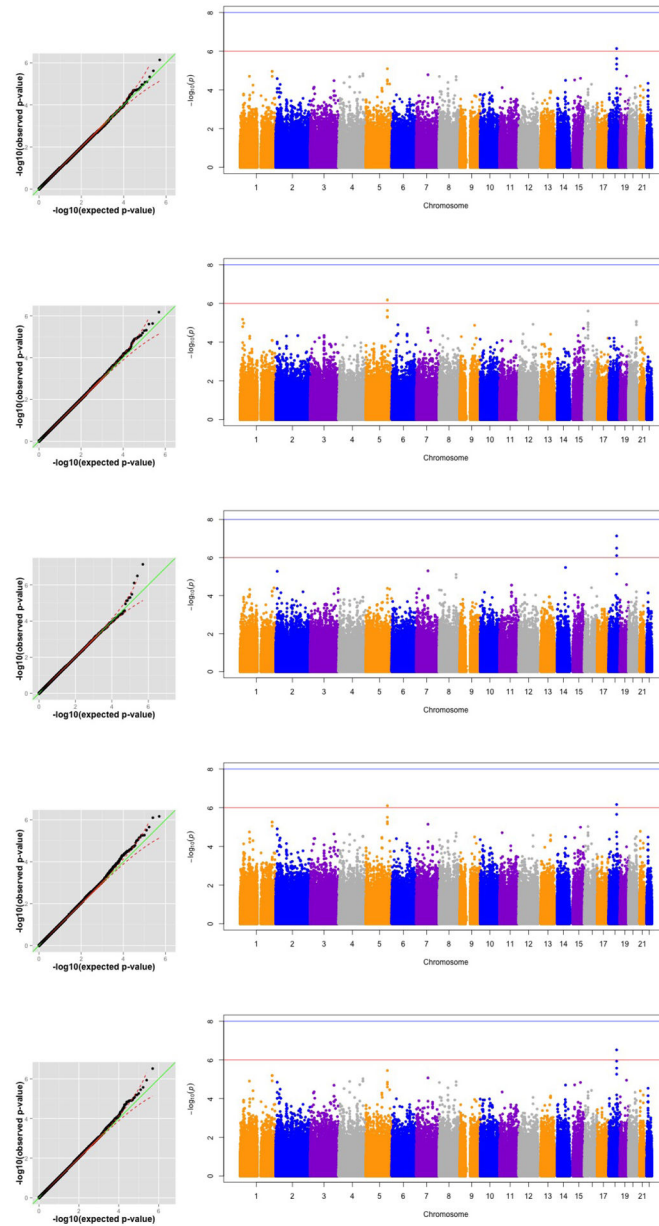


Figure 7. ADNI data analysis results: Q-Q and Manhattan plots of genome-wide association study (GWAS) of the left amygdala volume by using LReg (first row), LRegD (second row), SEE with $dp = 0.25$ (third row), and SPREG with $dp = 0.15$ (fourth row) and $dp = 0.35$ (fifth row).

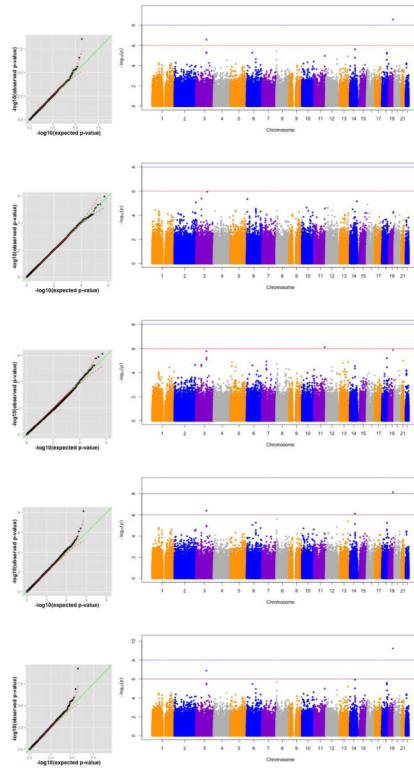


Figure 8. ADNI data analysis results: Q-Q and Manhattan plots of genome-wide association study (GWAS) of the right amygdala volume by using LReg (first row), LRegD (second row), SEE with $dp = 0.2$ (third row), and SPREG with $dp = 0.15$ (fourth row) and $dp = 0.35$ (fifth row).

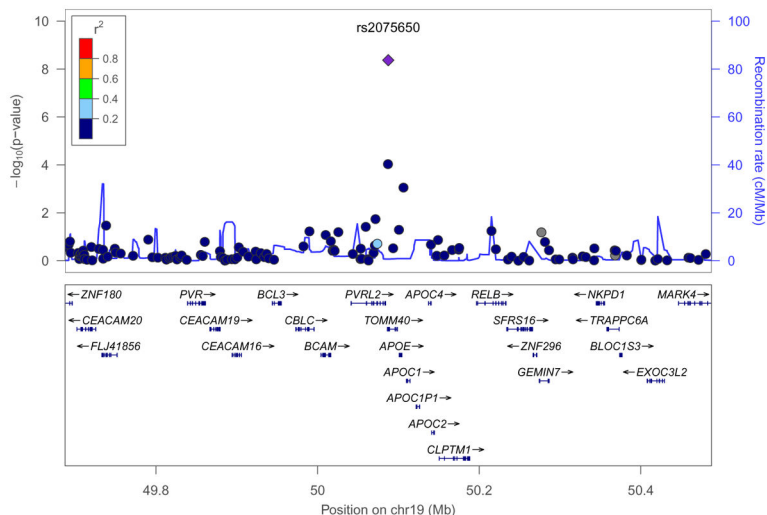


Figure 9. ADNI data analysis results: the LocusZoom plot of genome-wide association study (GWAS) of the right hippocampus volume by using SPREG showing ADNI associated region near the TOMM40 gene and intergenic region. Pairwise values of LD with the top SNP (rs2075650 in purple) was calculated using the HapMap CEU population. Physical positions are based on NCBI Build 36 of the human genome.

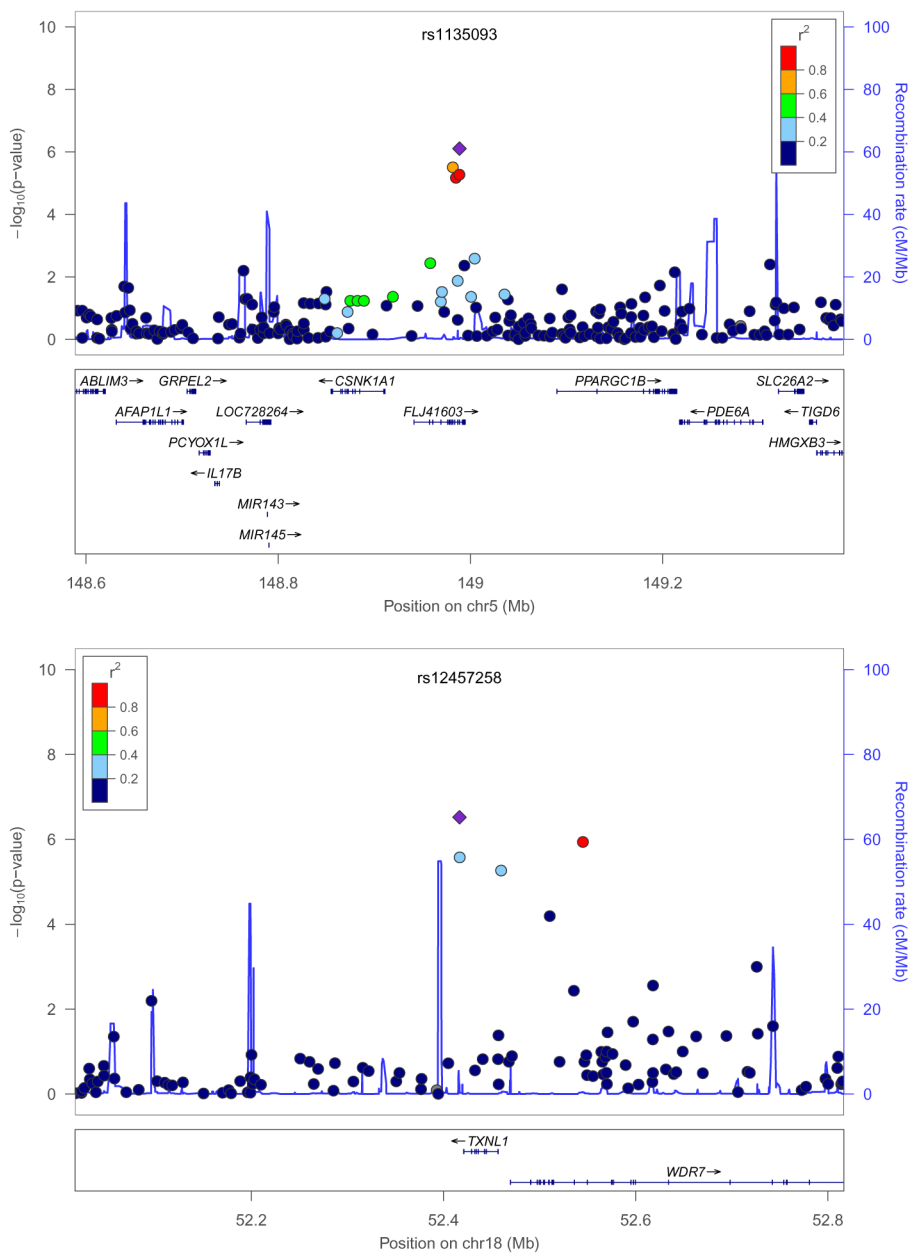


Figure 10. ADNI data analysis results: the LocusZoom plot of genome-wide association study (GWAS) of the left amygdala volume by using SPREG showing ADNI associated region near the FLJ41603 and TXNL1 genes. Pairwise values of LD with the top SNPs (rs1135093 and rs12457258 in purple) were calculated using the HapMap CEU population. Physical positions are based on NCBI Build 36 of the human genome.

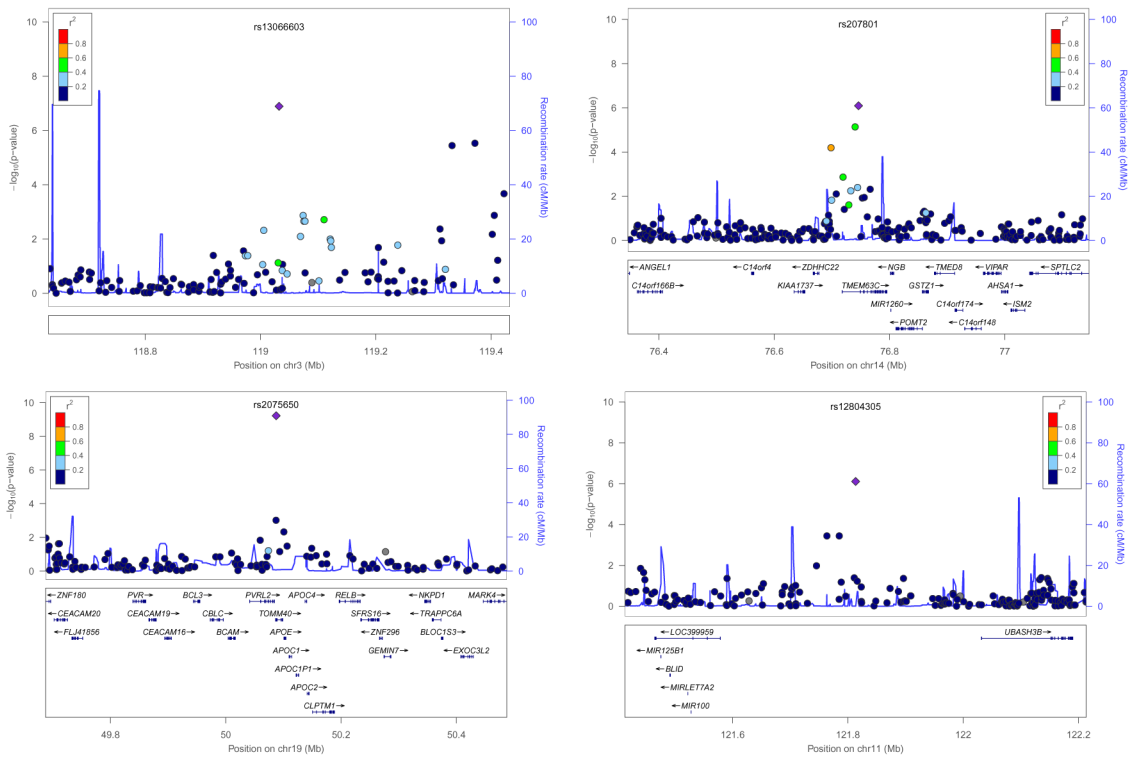


Figure 11. ADNI data analysis results: the LocusZoom plot of genome-wide association study (GWAS) of the right amygdala volume by using SPREG and SEE (the lower right) showing ADNI associated region near the TMEM63C, the TOMM40 gene and intergenic regions. Pairwise values of LD with the top SNPs (rs13066603, rs207801, rs2075650, and rs12804305 in purple) were calculated using the HapMap CEU population. Physical positions are based on NCBI Build 36 of the human genome.

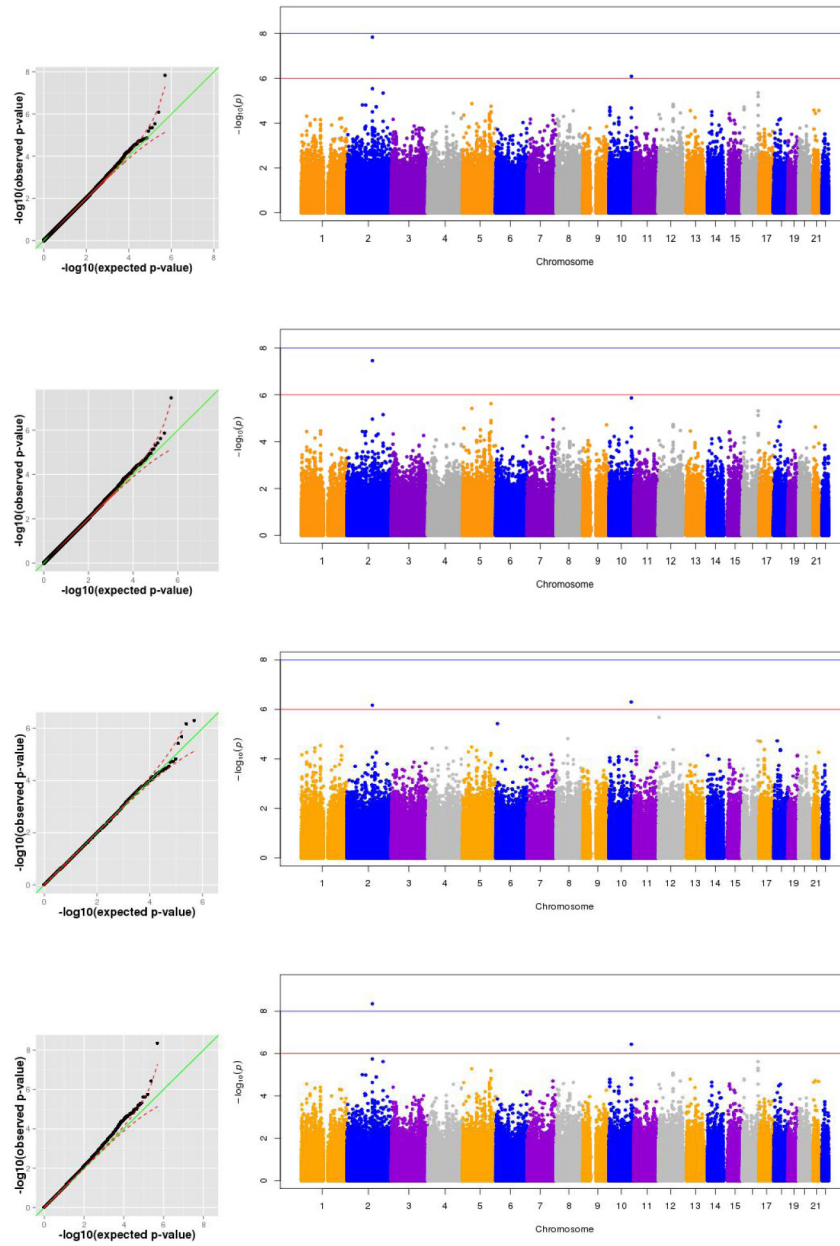


Figure 12. ADNI data analysis results: Q-Q and Manhattan plots of genome-wide association study (GWAS) of the volume of me.f-o.gy.R by using LReg (first row), LRegD (second row), SEE with $dp = 0.25$ (third row), and SPREG with $dp = 0.25$ (fourth row).

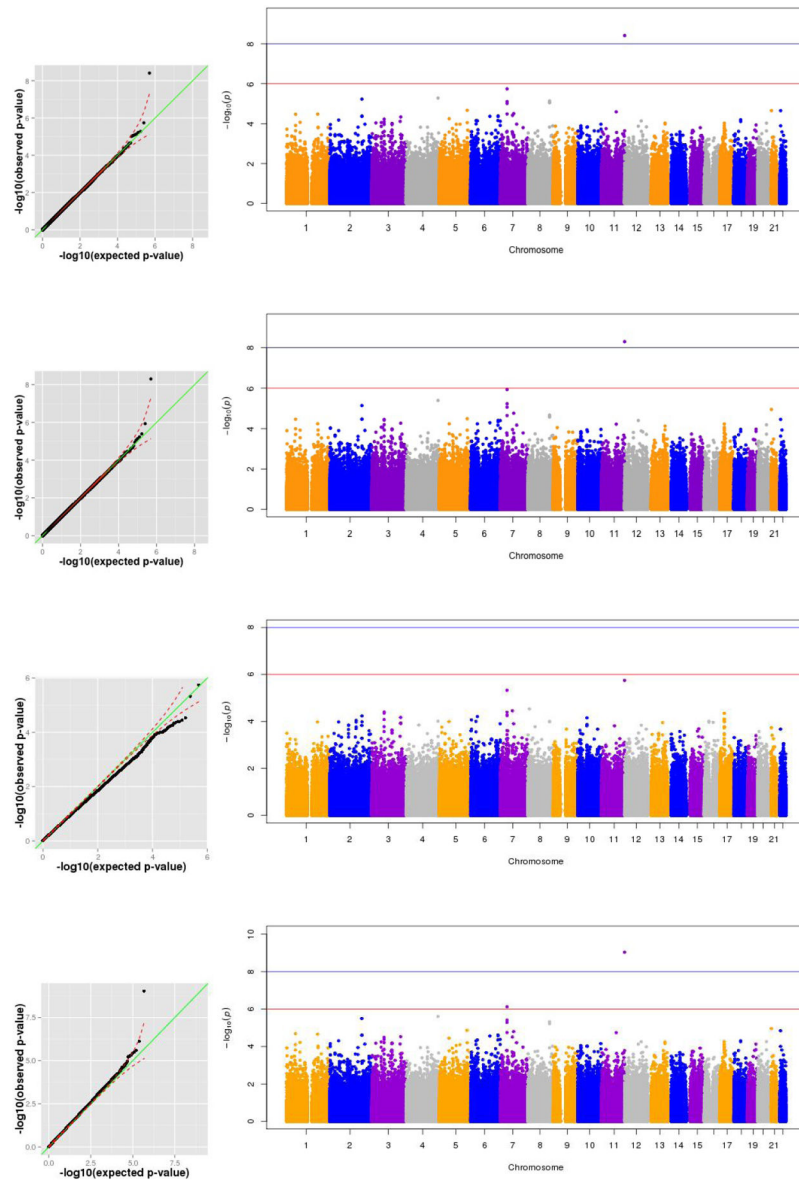


Figure 13. ADNI data analysis results: Q-Q and Manhattan plots of genome-wide association study (GWAS) of the volume of sup.f.gy.R by using LReg (first row), LRegD (second row), SEE with $dp = 0.25$ (third row), and SPREG with $dp = 0.25$ (fourth row).

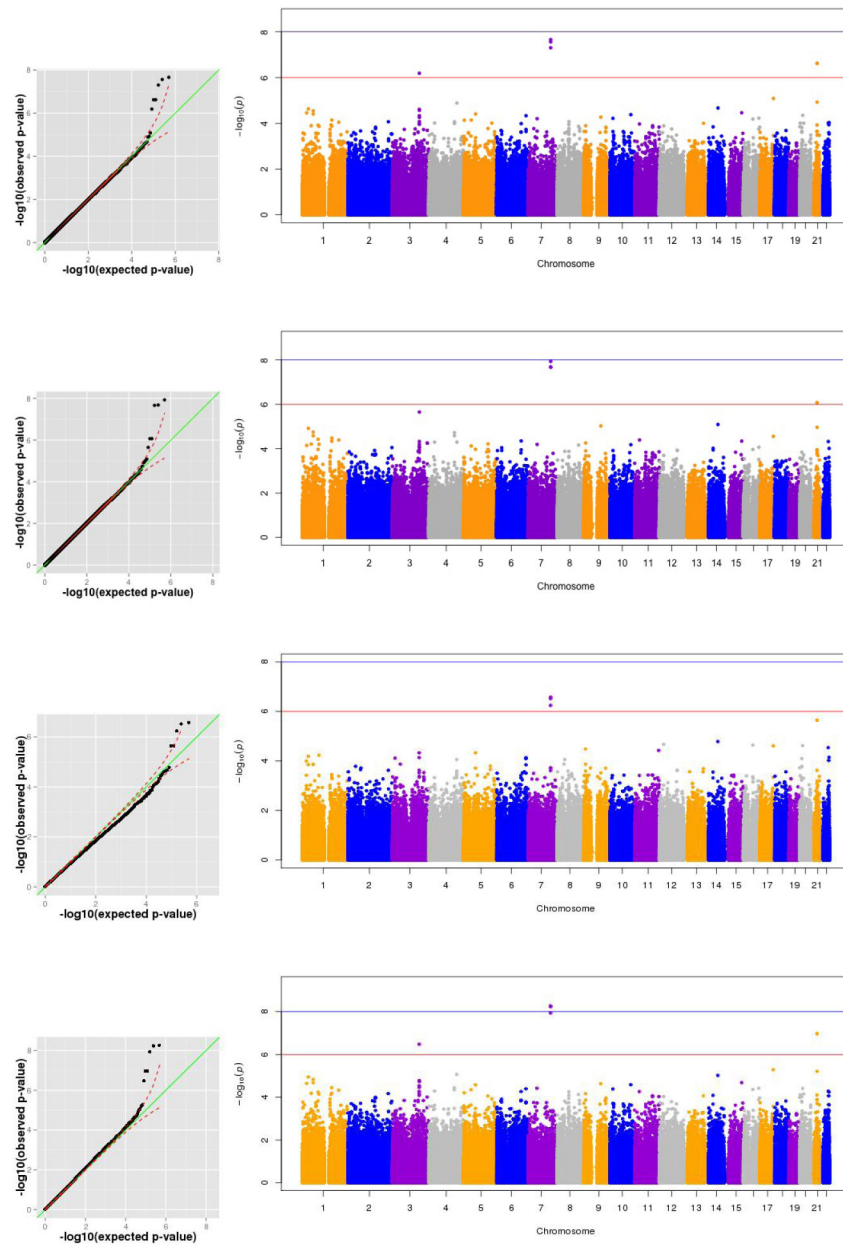


Figure 14. ADNI data analysis results: Q-Q and Manhattan plots of genome-wide association study (GWAS) of the volume of inf.f.gy.L by using LReg (first row), LRegD (second row), SEE with $dp = 0.25$ (third row), and SPREG with $dp = 0.25$ (fourth row).

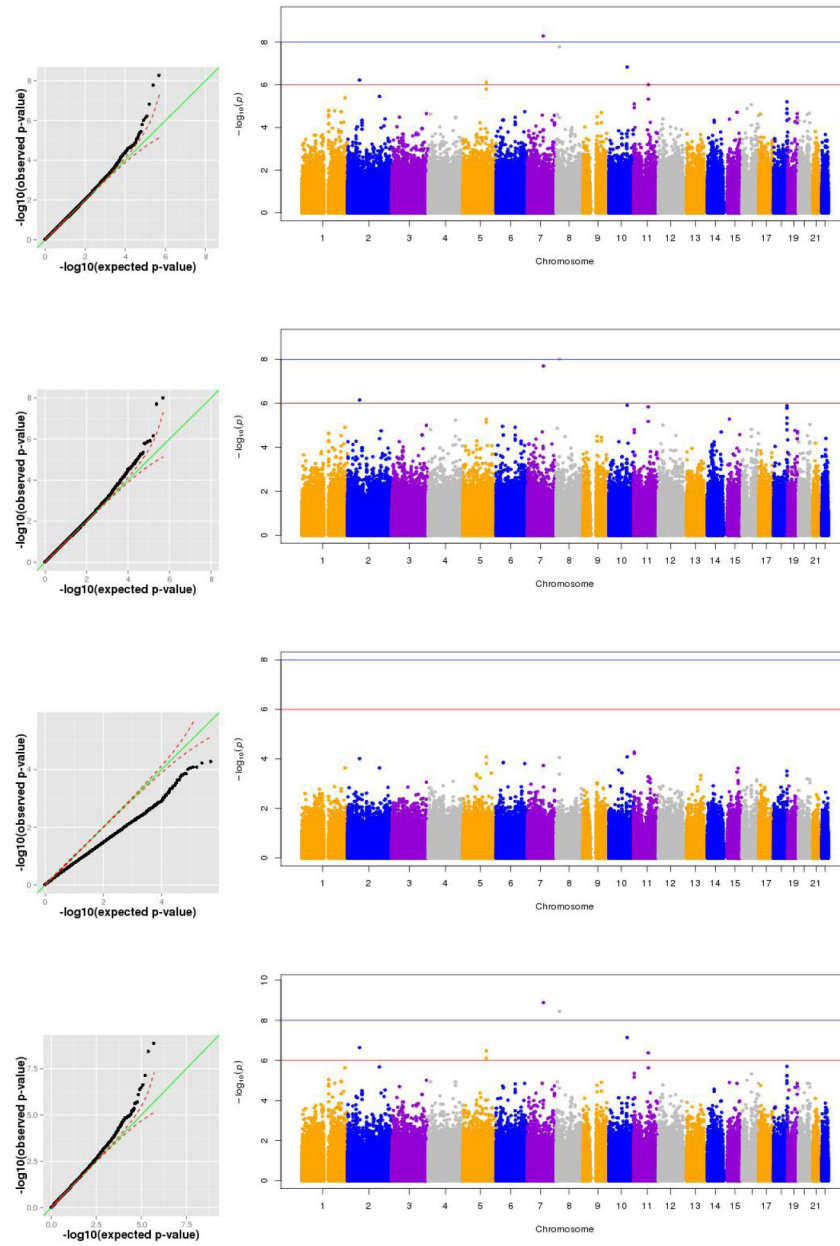


Figure 15. ADNI data analysis results: Q-Q and Manhattan plots of genome-wide association study (GWAS) of the volume of fornix.R by using LReg (first row), LRegD (second row), SEE with $dp = 0.25$ (third row), and SPREG with $dp = 0.25$ (fourth row).

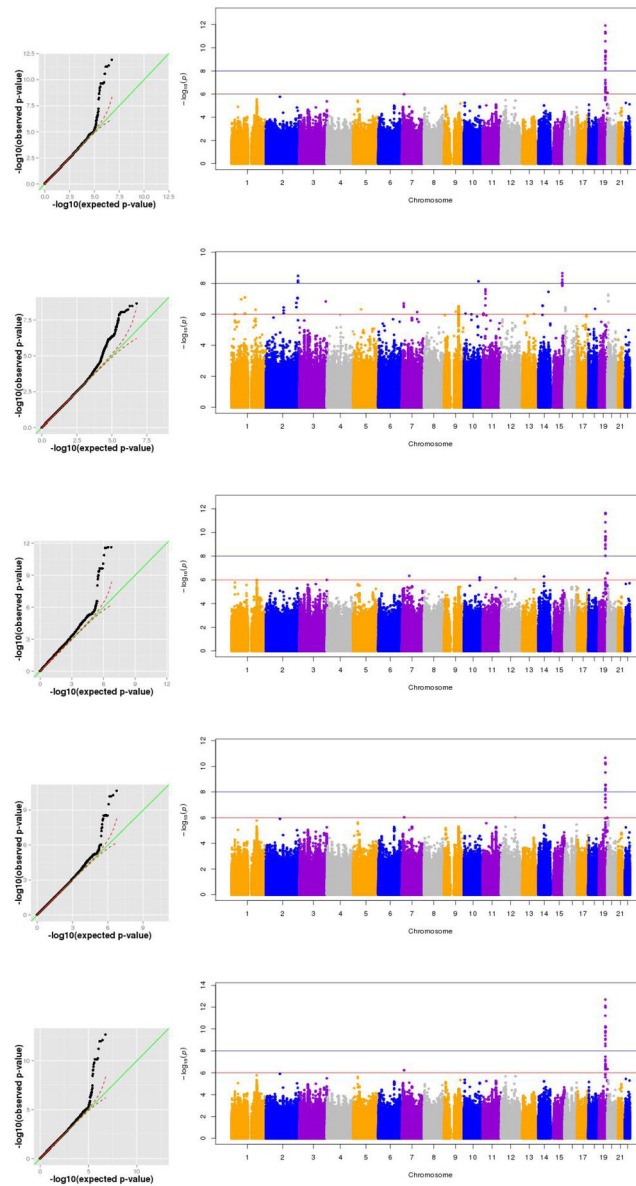


Figure 16. ADNI 1 and 2 data analysis results: Q-Q and Manhattan plots of genome-wide association study (GWAS) of the right hippocampus volume by using LReg (first row), LRegD (second row), SEE with $dp = 0.25$ (third row), and SPREG with $dp = 0.15$ (fourth row) and $dp = 0.35$ (fifth row).

Table 1

Simulation results: comparisons of Type I error rates for the analysis of SNP-secondary trait association in a case-control sample with 10% disease prevalence based on 100,000 simulations. OR_{DY} (OR_{DC}) is the odds ratio of the secondary trait (SNP) with the case-control status. α represents the nominal significance level.

	OR_{DY}	0.5				2			
		0	1.2	1.4	1.4	0	1.2	1.2	1.4
MAF = 0.1	$\alpha = 0.01$	0.01106	0.01757	0.02367	0.00945	0.00737	0.01461		
		0.00939	0.00778	0.01280	0.00979	0.01821	0.01833		
		0.00825	0.00863	0.01086	0.00753	0.00764	0.00877		
		0.01107	0.01107	0.01086	0.01199	0.01218	0.01117		
		0.01047	0.00936	0.00818	0.00988	0.01054	0.00706		
$\alpha = 0.05$		0.05414	0.07455	0.09200	0.04724	0.04072	0.06704		
		0.04811	0.04399	0.05756	0.04882	0.07656	0.07747		
		0.04414	0.04538	0.05242	0.04130	0.04251	0.04470		
		0.05314	0.05203	0.05302	0.05579	0.05598	0.05412		
		0.0511	0.04808	0.04439	0.04867	0.05141	0.04074		
MAF = 0.3	$\alpha = 0.01$	0.01608	0.05915	0.11076	0.00681	0.01641	0.04720		
		0.00943	0.00616	0.00877	0.00678	0.01246	0.02173		
		0.00581	0.00592	0.00862	0.00897	0.00683	0.00680		
		0.01095	0.01131	0.01117	0.01130	0.01144	0.01182		
		0.01076	0.00988	0.00664	0.00673	0.00762	0.00814		
$\alpha = 0.05$		0.07021	0.17971	0.28638	0.03997	0.07009	0.15600		
		0.04879	0.03735	0.04753	0.03938	0.06167	0.08904		
		0.03720	0.03787	0.04604	0.04624	0.03938	0.03857		
		0.05362	0.05384	0.05384	0.05616	0.05451	0.05399		
		0.05372	0.04971	0.03845	0.03873	0.04315	0.04454		
MAF = 0.45	$\alpha = 0.01$	0.00952	0.02211	0.05625	0.0094	0.01197	0.03185		
		0.00832	0.01266	0.02546	0.01075	0.02233	0.04506		
		0.00924	0.01176	0.01290	0.00760	0.00796	0.00769		
		0.01118	0.01147	0.01184	0.01250	0.01267	0.01190		

	0.5				2			
	0	1.2	1.4	1.4	0	1.2	1.4	1.4
OR_{DY}								
OR_{DG}								
SPREG	0.00877	0.00846	0.00848	0.00848	0.01022	0.0095	0.00917	0.00917
LReg	0.04843	0.08773	0.17027	0.17027	0.04870	0.05803	0.111499	0.111499
LRegD	0.04347	0.05872	0.09488	0.09488	0.05254	0.08822	0.14582	0.14582
LRegDG	0.04746	0.05679	0.05913	0.05913	0.04320	0.04426	0.04430	0.04430
SEE	0.05495	0.05394	0.05468	0.05468	0.05587	0.05755	0.05570	0.05570
SPREG	0.04576	0.04366	0.04496	0.04496	0.05104	0.04729	0.04709	0.04709

$\alpha = 0.05$

ADNI data analysis results: the SNP-ROI pairs with p -value $< 10^{-6}$ using the four methods in Section 2 for different disease prevalences.

Table 2

Method	dp	right hippocampus			left amygdala			right amygdala		
		rs2075650	rs1135093	rs12457258	rs13066603	rs207801	rs2075650	rs12804305		
LReg	/	1.52E-08	8.03E-06	7.26E-07	2.75E-07	2.44E-06	2.86E-09	1.22E-04		
LRegD	/	8.11E-03	6.62E-07	8.33E-04	2.92E-04	5.09E-05	3.36E-03	2.36E-02		
SEE	0.10	1.42E-03	5.17E-03	1.97E-04	3.32E-04	3.21E-03	5.02-03	1.99E-05		
	0.15	3.44E-05	1.28E-03	1.05E-05	2.53E-05	4.17E-04	9.25E-05	3.17E-06		
	0.20	7.40E-07	2.52E-04	6.99E-07	1.69E-06	4.48E-05	1.33E-06	7.84E-07		
	0.25	1.96E-08	4.14E-05	7.30E-08	1.18E-07	4.57E-06	2.22E-08	3.28E-07		
	0.30	8.38E-10	6.05E-06	1.42E-08	1.04E-08	5.21E-07	5.68E-10	2.42E-07		
	0.35	6.99E-11	8.65E-07	5.67E-09	1.38E-09	7.80E-08	2.77E-11	3.09E-07		
	0.40	1.30E-11	1.36E-07	4.55E-09	3.18E-10	1.76E-08	2.98E-12	6.40E-07		
SPREG	0.10	3.05E-07	4.39E-07	1.25E-06	8.08E-07	7.25E-07	2.63E-08	2.71E-04		
	0.15	6.49E-08	7.89E-07	6.83E-07	4.06E-07	8.08E-07	7.38E-09	1.73E-04		
	0.20	2.15E-08	1.31E-06	4.72E-07	2.57E-07	9.30E-07	2.94E-09	1.28E-04		
	0.25	9.79E-09	2.00E-06	3.73E-07	1.87E-07	1.05E-06	1.44E-09	1.04E-04		
	0.30	5.77E-09	2.78E-06	3.23E-07	1.50E-07	1.14E-06	8.40E-10	8.92E-05		
	0.35	4.26E-09	3.51E-06	3.01E-07	1.29E-07	1.19E-06	5.98E-10	8.15E-05		
	0.40	3.82E-09	4.09E-06	2.98E-07	1.17E-07	1.21E-06	5.25E-10	7.88E-05		

dp: disease prevalence

ADNI data analysis results: the numbers of SNP-ROI pairs with p -value $< 10^{-6}$ that are identified by SPREG and/or SEE but not by LReg and LRegD at disease prevalence = 0.25.

Table 3

ROI	LReg	LRegD	SEE	SPREG	ROI	LReg	LRegD	SEE	SPREG
lat.ve.L	0	0	0	1	me.f.gy.R	0	0	0	1
me.f.gy.L	0	0	0	1	ant.caps.R	0	0	0	1
psic.gy.L	0	0	1	0	inf.t.gy.R	0	0	1	0
inf.t.gy.L	0	0	1	1	ang.gyr.L	0	0	0	1
par.lb.WM.R	0	0	0	1	thalR	0	0	0	1
ling.gy.R	0	0	1	0					

ADNI data analysis results: the numbers of SNP-ROI pairs with p -value $< 10^{-6}$ at disease prevalence = 0.25, and the p -values of SPREG and/or SEE are smaller than those of LReg and LRegD.

Table 4

ROI	LReg	LRegD	SEE	SPREG	ROI	LReg	LRegD	SEE	SPREG
me.f.o.gy.R	2	2	2	2	par.lb.WM.L	1	1	1	3
mid.f.gy.R	1	1	0	1	sup.gy.R	2	2	2	3
insula.R	1	0	0	2	mid.t.gy.L	1	0	1	2
lat.f-o.gy.R	2	0	0	3	oc.lb.WM.L	2	2	0	3
lat.ve.R	1	0	1	1	inf.f.gy.R	1	0	0	1
sup.f.gy.R	1	1	0	2	me.f-o.gy.L	0	1	1	3
inf.f.gy.L	5	3	3	5	per.cort.R	2	1	0	4
f.lob.WMR	1	2	1	2	ent.cort.L	2	0	0	2
tmp.pl.R	3	0	0	4	inf.o.gy.R	1	0	1	2
subtha.nuc.R	1	1	1	2	sup.o.gy.L	1	1	0	2
unc.R	0	1	2	1	lat.Lo.t.gy.R	4	1	1	5
fornix.L	1	2	0	3	thalL	0	1	1	3
post.limb.R	1	1	0	1	parah.gy.R	2	0	0	3
caud.neuc.L	1	1	0	2	corp.col	1	1	0	2
sup.gy.L	2	2	0	3	sup.t.gy.R	3	0	4	3
sup.p.lb.L	1	0	0	2	cun.R	1	1	0	1
caud.neuc.R	2	2	0	2	occ.pol.L	0	1	0	1
cun.L	2	2	0	2	fornix.R	6	3	0	7
prec.L	1	0	0	1					