Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 843563, 12 pages
http://dx.doi.org/10.1155/2013/843563

Research Article

Robust Joint Analysis with Data Fusion in Two-Stage Quantitative Trait Genome-Wide Association Studies

Dong-Dong Pan,[1] Wen-Jun Xiong,[2] Ji-Yuan Zhou,[3] Ying Pan,[4] Guo-Li Zhou,[5] and Wing-Kam Fung[6]

[1] *Department of Statistics, Yunnan University, Kunming 650091, China*
[2] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*
[3] *Department of Biostatistics, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou 510515, China*
[4] *Department of Biology, Nanjing University, Nanjing 210093, China*
[5] *College of Mathematics and Statistics, Chongqing University, Chongqing 40044, China*
[6] *Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong*

Correspondence should be addressed to Dong-Dong Pan; ddpan@ynu.edu.cn and Wing-Kam Fung; wingfung@hku.hk

Received 6 July 2013; Accepted 29 July 2013

Academic Editor: Qizhai Li

Genome-wide association studies (GWASs) in identifying the disease-associated genetic variants have been proved to be a great pioneering work. Two-stage design and analysis are often adopted in GWASs. Considering the genetic model uncertainty, many robust procedures have been proposed and applied in GWASs. However, the existing approaches mostly focused on binary traits, and few work has been done on continuous (quantitative) traits, since the statistical significance of these robust tests is difficult to calculate. In this paper, we develop a powerful F-statistic-based robust joint analysis method for quantitative traits using the combined raw data from both stages in the framework of two-staged GWASs. Explicit expressions are obtained to calculate the statistical significance and power. We show using simulations that the proposed method is substantially more robust than the F-test based on the additive model when the underlying genetic model is unknown. An example for rheumatic arthritis (RA) is used for illustration.

1. Introduction

Genome-wide association studies (GWASs) have identified a large number of genomic regions (especially single-nucleotide polymorphisms (SNPs)) with a wide variety of complex traits/diseases. In a GWAS, two most common types of data, qualitative (or binary) and quantitative (or continuous) traits, are analyzed and two contentious points are often faced; one is how to construct the test statistic considering the genetic model uncertainty and the other is how to evaluate the statistical significance for controlling the false positive rates efficiently (e.g., [1, 2]). Considering these issues, a lot of work has been done on the binary trait in the past 10 years (e.g., [3–7]). Computer algorithms have also been developed to calculated the significance level of robust tests in GWASs, taking into account the genetic model uncertainty [8]. However, few work has been done on continuous traits, only recently So and Sham [9] proposed a MAX3 based on score test statistics, and Li et al. [10] gave a MAX3 based on F-test statistics. Note that these tests just focus on single-marker analysis in one-stage analysis.

Although the costs of whole-genome genotyping are decreasing with the high-throughput biological technology, the total costs for a GWAS are still very expensive due to the thousands of sampling units and huge amounts of single-nucleotide polymorphisms. In order to save the costs, the two-stage design and the corresponding statistical analysis where all the SNPs are genotyped in Stage 1 on a portion of the samples and the promising SNPs with small P-values (e.g., <0.001) based on some efficient tests are further screened on the remaining subjects, are often adopted in practice (e.g., [11–15]).

In genetic association studies, especially GWASs, genetic markers are routinely tested under the assumption of additive

effects. Although convenient to use, those tests are optimal only when the true underlying genetic model is additive so that they are not robust against the genetic model misspecification. To our best knowledge, few work has been done on the two-stage joint analysis for quantitative trait GWASs allowing for genetic model uncertainty. Here, we attempt to develop a joint analysis method with data fusion in the two-stage design using F-statistic, since F-test is commonly employed from the linear regression model for quantitative trait, and Li et al. [10] show that MAX3 based on F-statistics is more powerful than So and Sham's method by extensively numerical simulation.

The content of this paper is organized as follows. In Section 2, we give some notations and the proposed robust joint test statistics. Further, we derive the asymptotic distribution of the test statistics under the null and the alternative hypotheses. In Section 3, we show that the proposed joint analysis method is substantially more robust than the additive-model-based F-test from the numerical results of power comparison when the real genetic model is unknown. After that, an illustrative example for rheumatic arthritis (RA) is presented. Finally, we give some discussion of this paper in Section 4.

2. Methods

2.1. Notations. Assume that n individuals are randomly selected to be genotyped in a two-staged GWAS for a certain quantitative trait and that π is the sampling proportion in Stage 1. Let $n_1 = n\pi$ and $n_2 = n(1-\pi)$ be the sample sizes for Stages 1 and 2, respectively. Consider a biallelic marker with two alleles G and g. Without loss of generality, we assume that G is the minor or high-risk allele. We suppose that the total m SNPs are genotyped on the samples of Stage 1, and SNPs with P-values less than γ in Stage 1 will be further genotyped and tested in Stage 2. Let the significance level be α, and then the genome-wide significance level per SNP is α/m with the Bonferroni adjustments. Let $\mathbf{Y}_1 = (y_1, y_2, \ldots, y_{n_1})'$ and $\mathbf{Y}_2 = (y_{n_1+1}, y_{n_1+2}, \ldots, y_n)'$ be the observed quantitative outcome vectors for Stage 1 and Stage 2, respectively. Without loss of generality, we assume that the first n_{10} individuals in Stage 1 have the genotype gg, the second n_{11} individuals in Stage 1 have the genotype Gg, and the last n_{12} subjects in Stage 1 possess the genotype GG. Similarly, the first n_{20} subjects in Stage 2 have the genotype gg, the second n_{21} individuals in Stage 2 have the genotype Gg, and the last n_{22} subjects in Stage 2 possess the genotype GG. Let $\mathbf{0}_k = (0, 0, \ldots, 0)'_{k\times 1}$ and $\mathbf{1}_k = (1, 1, \ldots, 1)'_{k\times 1}$, and let $\mathbf{O}_{k\times j}$ be the $k \times j$ matrix with all its entries being zero and $\mathbf{I}_n$ be the $n \times n$ identity matrix.

2.2. F-Statistic-Based Robust Joint Analysis. We firstly briefly introduce F-statistic-based MAX3 by Li et al. [10] just using the data from Stage 1. Consider the following linear regression model:

$$y_i = \beta_0 + g_i\beta_1 + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \ i = 1, 2, \ldots, n_1, \quad (1)$$

where β_0 is the nuisance parameter for the intercept, β_1 is the parameter of interest for genetic effect, and g_i is the genotype value, which takes 0, 1, or 2 corresponding to the count of G at a marker locus for the ith subject, $i = 1, 2, \ldots, n_1$. The hypotheses of interest are

$$H_0 : \beta_1 = 0 \longleftrightarrow H_1 : \beta_1 \neq 0. \quad (2)$$

The variable g_i in the previously stated equation is coded differently for the three common genetic models. Let $\mathbf{X}_{1R} = (\mathbf{1}_{n_1}, \mathbf{G}_{1R})$, $\mathbf{X}_{1A} = (\mathbf{1}_{n_1}, \mathbf{G}_{1A})$, and $\mathbf{X}_{1D} = (\mathbf{1}_{n_1}, \mathbf{G}_{1D})$ be the design matrices under three commonly used genetic models, where $\mathbf{G}_{1R} = (\mathbf{0}'_{n_{10}+n_{11}}, \mathbf{1}'_{n_{12}})'$ corresponds to the recessive model, $\mathbf{G}_{1A} = (g_1, g_2, \ldots, g_{n_1})'$ corresponds to the additive model, and $\mathbf{G}_{1D} = (\mathbf{0}'_{n_{10}}, \mathbf{1}'_{n_{11}+n_{12}})'$ is for the dominant model. Denote $\mathbf{X}_1 = (\mathbf{1}_{n_1}, \mathbf{x}_{11}, \mathbf{x}_{12})$, where $\mathbf{x}_{11} = (\mathbf{0}'_{n_{10}}, \mathbf{1}'_{n_{11}}, \mathbf{0}'_{n_{12}})'$ and $\mathbf{x}_{12} = (\mathbf{0}'_{n_{10}}, \mathbf{0}'_{n_{11}}, \mathbf{1}'_{n_{12}})'$. The modified F-test statistics under the recessive, additive, and dominant models for Stage 1 are given by

$$\begin{aligned}
F_1^R &= \frac{\mathbf{Y}_1'\left[\mathbf{X}_{1R}\left(\mathbf{X}_{1R}'\mathbf{X}_{1R}\right)^{-1}\mathbf{X}_{1R}' - \mathbf{1}_{n_1}\left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1}\mathbf{1}_{n_1}'\right]\mathbf{Y}_1}{\mathbf{Y}_1'\left[\mathbf{I}_{n_1} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{Y}_1/\left(n_1-3\right)} \\
&= \frac{\left(Z_1^R\right)^2}{\mathrm{RSS}_1/\left(n_1-3\right)}, \\
F_1^A &= \frac{\mathbf{Y}_1'\left[\mathbf{X}_{1A}\left(\mathbf{X}_{1A}'\mathbf{X}_{1A}\right)^{-1}\mathbf{X}_{1A}' - \mathbf{1}_{n_1}\left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1}\mathbf{1}_{n_1}'\right]\mathbf{Y}_1}{\mathbf{Y}_1'\left[\mathbf{I}_{n_1} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{Y}_1/\left(n_1-3\right)} \\
&= \frac{\left(Z_1^A\right)^2}{\mathrm{RSS}_1/\left(n_1-3\right)}, \\
F_1^D &= \frac{\mathbf{Y}_1'\left[\mathbf{X}_{1D}\left(\mathbf{X}_{1D}'\mathbf{X}_{1D}\right)^{-1}\mathbf{X}_{1D}' - \mathbf{1}_{n_1}\left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1}\mathbf{1}_{n_1}'\right]\mathbf{Y}_1}{\mathbf{Y}_1'\left[\mathbf{I}_{n_1} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{Y}_1/\left(n_1-3\right)} \\
&= \frac{\left(Z_1^D\right)^2}{\mathrm{RSS}_1/\left(n_1-3\right)},
\end{aligned} \quad (3)$$

where

$$\begin{aligned}
Z_1^R &= \sqrt{\frac{\left(n_{10}+n_{11}\right)n_{12}}{n_1}}\left(\overline{y}_{n_{10}+n_{11}} - \overline{y}_{n_{12}}\right), \\
Z_1^D &= \sqrt{\frac{n_{10}\left(n_{11}+n_{12}\right)}{n_1}}\left(\overline{y}_{n_{10}} - \overline{y}_{n_{11}+n_{12}}\right), \\
Z_1^A &= \left(n_{10}\left(n_{11}+2n_{12}\right)\overline{y}_{n_{10}} - n_{11}\left(n_{10}-n_{12}\right)\overline{y}_{n_{11}}\right. \\
&\quad \left. - n_{12}\left(2n_{10}+n_{11}\right)\overline{y}_{n_{12}}\right) \\
&\quad \times \left(\sqrt{n_1\left[n_{10}\left(n_{11}+4n_{12}\right)+n_{11}n_{12}\right]}\right)^{-1}, \\
\overline{y}_{n_{10}} &= \frac{1}{n_{10}}\sum_{j=1}^{n_{10}} y_j,
\end{aligned}$$

$$\overline{y}_{n_{11}} = \frac{1}{n_{11}} \sum_{j=n_{10}+1}^{n_{10}+n_{11}} y_j,$$

$$\overline{y}_{n_{12}} = \frac{1}{n_{12}} \sum_{j=n_{10}+n_{11}+1}^{n_1} y_j,$$

$$\overline{y}_{n_{10}+n_{11}} = \frac{1}{n_{10}+n_{11}} \sum_{j=1}^{n_{10}+n_{11}} y_j,$$

$$\overline{y}_{n_{11}+n_{12}} = \frac{1}{n_{11}+n_{12}} \sum_{j=n_{10}+1}^{n_1} y_j. \tag{4}$$

The robust test statistic in Stage 1 is

$$F_1^{\mathrm{MAX}} = \max\left\{F_1^R, F_1^A, F_1^D\right\}. \tag{5}$$

We now give the proposed robust joint analysis. In the framework of two-stage design GWAS of quantitative traits, the SNPs with P-values less than γ will be genotyped on the remaining n_2 subjects in Stage 2. Following the previous notation for Stage 1, corresponding to the recessive, additive, and dominant models, the genotype data in Stage 2 are denoted by $\mathbf{G}_{2R} = (\mathbf{0}'_{n_{20}+n_{21}}, \mathbf{1}'_{n_{22}})'$, $\mathbf{G}_{2A} = (g_{n_1+1}, g_{n_1+2}, \ldots, g_n)'$, and $\mathbf{G}_{2D} = (\mathbf{0}'_{n_{20}}, \mathbf{1}'_{n_{21}+n_{22}})'$, respectively, and the design matrices are $\mathbf{X}_{2R} = (\mathbf{1}_{n_2}, \mathbf{G}_{2R})$, $\mathbf{X}_{2A} = (\mathbf{1}_{n_2}, \mathbf{G}_{2A})$, and $\mathbf{X}_{2D} = (\mathbf{1}_{n_2}, \mathbf{G}_{2D})$, respectively. Denote $\mathbf{X}_2 = (\mathbf{1}_{n_2}, \mathbf{x}_{21}, \mathbf{x}_{22})$, where $\mathbf{x}_{21} = (\mathbf{0}'_{n_{20}}, \mathbf{1}'_{n_{21}}, \mathbf{0}'_{n_{22}})'$ and $\mathbf{x}_{22} = (\mathbf{0}'_{n_{20}}, \mathbf{0}'_{n_{21}}, \mathbf{1}'_{n_{22}})'$. Then, we can obtain three modified F-test statistics under the recessive, additive, and dominant models for Stage 2 similarly, and denote them by F_2^R, F_2^A, and F_2^D. Let $\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2)'$, $\mathbf{G}_R = (\mathbf{G}'_{1R}, \mathbf{G}'_{2R})'$, $\mathbf{G}_A = (\mathbf{G}'_{1A}, \mathbf{G}'_{2A})'$, and $\mathbf{G}_D = (\mathbf{G}'_{1D}, \mathbf{G}'_{2D})'$. Denote $N_0 = n_{10} + n_{20}$, $N_1 = n_{11} + n_{21}$, and $N_2 = n_{12} + n_{22}$ for the combined sample sizes from two stages, corresponding to three genotypes. Then the proposed F-test statistics under three genetic models on the basis of the combined data are as follows:

$$\begin{aligned}
F_J^R &= \frac{\mathbf{Y}'\left[\mathbf{X}_R\left(\mathbf{X}'_R\mathbf{X}_R\right)^{-1}\mathbf{X}'_R - \mathbf{1}_n\left(\mathbf{1}'_n\mathbf{1}_n\right)^{-1}\mathbf{1}'_n\right]\mathbf{Y}}{\mathbf{Y}'\left[\mathbf{I}_n - \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'\right]\mathbf{Y}/(n-6)} \\
&= \frac{\left(Z_J^R\right)^2}{\mathrm{RSS}_J/(n-6)}, \\
F_J^A &= \frac{\mathbf{Y}'\left[\mathbf{X}_A\left(\mathbf{X}'_A\mathbf{X}_A\right)^{-1}\mathbf{X}'_A - \mathbf{1}_n\left(\mathbf{1}'_n\mathbf{1}_n\right)^{-1}\mathbf{1}'_n\right]\mathbf{Y}}{\mathbf{Y}'\left[\mathbf{I}_n - \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'\right]\mathbf{Y}/(n-6)} \\
&= \frac{\left(Z_J^A\right)^2}{\mathrm{RSS}_J/(n-6)}, \\
F_J^D &= \frac{\mathbf{Y}'\left[\mathbf{X}_D\left(\mathbf{X}'_D\mathbf{X}_D\right)^{-1}\mathbf{X}'_D - \mathbf{1}_n\left(\mathbf{1}'_n\mathbf{1}_n\right)^{-1}\mathbf{1}'_n\right]\mathbf{Y}}{\mathbf{Y}'\left[\mathbf{I}_n - \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'\right]\mathbf{Y}/(n-6)} \\
&= \frac{\left(Z_J^D\right)^2}{\mathrm{RSS}_J/(n-6)},
\end{aligned} \tag{6}$$

where $\mathbf{X}_R = (\mathbf{1}_n, \mathbf{G}_R)$, $\mathbf{X}_A = (\mathbf{1}_n, \mathbf{G}_A)$, $\mathbf{X}_D = (\mathbf{1}_n, \mathbf{G}_D)$, and $\mathbf{W} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{o}_{n_1\times 3} \\ \mathbf{o}_{n_2\times 3} & \mathbf{X}_2 \end{pmatrix}$,

$$\begin{aligned}
Z_J^R &= \sqrt{\frac{(N_0+N_1)N_2}{n}}\left(\overline{y}_{01} - \overline{y}_2\right), \\
Z_J^D &= \sqrt{\frac{N_0(N_1+N_2)}{n}}\left(\overline{y}_0 - \overline{y}_{12}\right), \\
Z_J^A &= \left(N_0(N_1+2N_2)\overline{y}_0 - N_1(N_0-N_2)\overline{y}_1\right. \\
&\quad \left. -N_2(2N_0+N_1)\overline{y}_2\right) \\
&\quad \times\left(\sqrt{n\left[N_0(N_1+4N_2)+N_1N_2\right]}\right)^{-1}, \\
\overline{y}_0 &= \frac{1}{N_0}\left(\sum_{j=1}^{n_{10}} y_j + \sum_{j=n_1+1}^{n_1+n_{20}} y_j\right), \\
\overline{y}_1 &= \frac{1}{N_1}\left(\sum_{j=n_{10}+1}^{n_{10}+n_{11}} y_j + \sum_{j=n_1+n_{20}+1}^{n_1+n_{20}+n_{21}} y_j\right), \\
\overline{y}_2 &= \frac{1}{N_2}\left(\sum_{j=n_{10}+n_{11}+1}^{n_1} y_j + \sum_{j=n_1+n_{20}+n_{21}+1}^{n} y_j\right), \\
\overline{y}_{01} &= \frac{N_0\overline{y}_0 + N_1\overline{y}_1}{N_0+N_1}, \\
\overline{y}_{12} &= \frac{N_1\overline{y}_1 + N_2\overline{y}_2}{N_1+N_2}.
\end{aligned} \tag{7}$$

Furthermore, we propose the joint testing statistic as

$$F_J^{\mathrm{MAX}} = \max\left\{F_J^R, F_J^A, F_J^D\right\}. \tag{8}$$

In order to calculate the power of the proposed joint analysis, we have to get the thresholds, which is determined by the significance level. Denote the threshold for choosing the promising SNPs in Stage 1 by u_1, which is the solution of

$$\Pr_{H_0}\left(F_1^{\mathrm{MAX}} > u_1\right) = \gamma. \tag{9}$$

Since the genome-wide significance level is α/m, in order to control the false positive rate, we have

$$\Pr_{H_0}\left(F_1^{\mathrm{MAX}} > u_1, F_J^{\mathrm{MAX}} > u_J\right) = \alpha/m, \tag{10}$$

where u_J is the cut-off point for the joint statistic. Once we have u_1 and u_J, the power is calculated by

$$\Pr_{H_1}\left(F_1^{\mathrm{MAX}} > u_1, F_J^{\mathrm{MAX}} > u_J\right). \tag{11}$$

We now give the detail to calculate the cut-off point and power above. The left side of (10) can be further expressed as

$$\begin{aligned}
&\Pr_{H_0}\left(F_1^{\mathrm{MAX}} > u_1, F_J^{\mathrm{MAX}} > u_J\right) \\
&\quad = 1 - \Pr_{H_0}\left(F_1^{\mathrm{MAX}} \le u_1\right) - \Pr_{H_0}\left(F_J^{\mathrm{MAX}} \le u_J\right) \\
&\qquad + \Pr_{H_0}\left(F_1^{\mathrm{MAX}} \le u_1, F_J^{\mathrm{MAX}} \le u_J\right).
\end{aligned} \tag{12}$$

For controlling the type I error rate and calculating the power, we need to know the distribution or the asymptotic distribution of $(Z_1^R, Z_1^A, Z_1^D, Z_J^R, Z_J^A, Z_J^D, \mathrm{RSS}_1, \mathrm{RRS}_J)'$ under both H_0 and H_1.

Note that whether H_0 or H_1 holds, RSS_1 and RSS_J and $(Z_1^R, Z_1^A, Z_1^D, Z_J^R, Z_J^A, Z_J^D)'$ are mutually independent (the proof is given in Appendix A). Denote the correlation matrix of $(Z_1^R, Z_1^A, Z_1^D)'$ by $\mathbf{V}_1 = (v_{kl})_{3\times 3}$, whose entries are $v_{11} = v_{22} = v_{33} = 1$, $v_{12} = v_{21} = \sqrt{n_{12}}\,(2n_{10} + n_{11})/\sqrt{(n_{10}+n_{11})[n_{10}(n_{11}+4n_{12})+n_{11}n_{12}]}$, $v_{13} = v_{31} = \sqrt{n_{10}n_{12}}/\sqrt{(n_{10}+n_{11})(n_{11}+n_{12})}$, and $v_{23} = v_{32} = \sqrt{n_{10}}\,(2n_{12} + n_{11})/\sqrt{(n_{11}+n_{12})[n_{10}(n_{11}+4n_{12})+n_{11}n_{12}]}$, respectively. Similarly, let $\mathbf{V}_J = (v_{kl}^*)_{3\times 3}$ be the correlation matrix of $(Z_J^R, Z_J^A, Z_J^D)'$ with $v_{11}^* = v_{22}^* = v_{33}^* = 1$, $v_{12}^* = v_{21}^* = \sqrt{N_2}(2N_0 + N_1)/\sqrt{(N_0+N_1)[N_0(N_1+4N_2)+N_1N_2]}$, $v_{13}^* = v_{31}^* = \sqrt{N_0N_2}/\sqrt{(N_0+N_1)(N_1+N_2)}$, and $v_{23}^* = v_{32}^* = \sqrt{N_0}(2N_2+ N_1)/\sqrt{(N_1+N_2)[N_0(N_1+4N_2)+N_1N_2]}$. Then, we can derive that $\mathrm{RRS}_1/\sigma^2 \sim \chi^2_{n_1-3}$, $\mathrm{RSS}_J/\sigma^2 \sim \chi^2_{n-6}$, and

$$\left(Z_1^R, Z_1^A, Z_1^D, Z_J^R, Z_J^A, Z_J^D\right)'\Big|_{H_0} \sim N_6\left(\mathbf{0}_6, \sigma^2\begin{pmatrix}\mathbf{V}_1 & \boldsymbol{\rho}\\ \boldsymbol{\rho}' & \mathbf{V}_J\end{pmatrix}\right), \tag{13}$$

where $\boldsymbol{\rho} = (\rho_{kl})_{3\times 3}$ is the correlation matrix between $(Z_1^R, Z_1^A, Z_1^D)'$ and $(Z_J^R, Z_J^A, Z_J^D)'$, with

$$\begin{aligned}
\rho_{11} &= \mathrm{Corr}\left(Z_1^R, Z_J^R\right) = \sqrt{\frac{n\left(n_{10}+n_{11}\right)n_{12}}{n_1\left(N_0+N_1\right)N_2}},\\
\rho_{12} &= \mathrm{Corr}\left(Z_1^R, Z_J^A\right)\\
&= \frac{\sqrt{nn_{12}}\left(2n_{10}+n_{11}\right)}{\sqrt{n_1\left(n_{10}+n_{11}\right)\left[N_0\left(N_1+4N_2\right)+N_1N_2\right]}},\\
\rho_{13} &= \mathrm{Corr}\left(Z_1^R, Z_J^D\right) = \frac{\sqrt{nn_{12}}n_{10}}{\sqrt{n_1\left(n_{10}+n_{11}\right)N_0\left(N_1+N_2\right)}},\\
\rho_{21} &= \mathrm{Corr}\left(Z_1^A, Z_J^R\right)\\
&= \frac{\sqrt{n}n_{12}\left(2n_{10}+n_{11}\right)}{\sqrt{n_1\left[n_{10}\left(n_{11}+4n_{12}\right)+n_{11}n_{12}\right]\left(N_0+N_1\right)N_2}},\\
\rho_{22} &= \mathrm{Corr}\left(Z_1^A, Z_J^A\right) = \sqrt{\frac{n\left[n_{10}\left(n_{11}+4n_{12}\right)+n_{11}n_{12}\right]}{n_1\left[N_0\left(N_1+4N_2\right)+N_1N_2\right]}},\\
\rho_{23} &= \mathrm{Corr}\left(Z_1^A, Z_J^D\right)\\
&= \frac{\sqrt{n}n_{10}\left(n_{11}+2n_{12}\right)}{\sqrt{n_1\left[n_{10}\left(n_{11}+4n_{12}\right)+n_{11}n_{12}\right]N_0\left(N_1+N_2\right)}},\\
\rho_{31} &= \mathrm{Corr}\left(Z_1^D, Z_J^R\right) = \frac{\sqrt{nn_{10}}n_{12}}{\sqrt{n_1\left(n_{11}+n_{12}\right)\left(N_0+N_1\right)N_2}},\\
\rho_{32} &= \mathrm{Corr}\left(Z_1^D, Z_J^A\right)\\
&= \frac{\sqrt{nn_{10}}\left(n_{11}+2n_{12}\right)}{\sqrt{n_1\left(n_{11}+n_{12}\right)\left[N_0\left(N_1+4N_2\right)+N_1N_2\right]}},\\
\rho_{33} &= \mathrm{Corr}\left(Z_1^D, Z_J^D\right) = \sqrt{\frac{nn_{10}\left(n_{11}+n_{12}\right)}{n_1N_0\left(N_1+N_2\right)}}.
\end{aligned} \tag{14}$$

Under H_1, for a given odds ratio $\mathrm{OR} = \exp(\beta_1)$ for subjects with two copies of risk allele corresponding to recessive model or one copy of risk allele corresponding to additive or dominant models, we have the following:

(i) when the true genetic model is recessive,

$$\left(Z_1^R, Z_1^A, Z_1^D, Z_J^R, Z_J^A, Z_J^D\right)'\Big|_{H_1} \sim N_6\left(\boldsymbol{\mu}^R, \sigma^2\begin{pmatrix}\mathbf{V}_1 & \boldsymbol{\rho}\\ \boldsymbol{\rho}' & \mathbf{V}_J\end{pmatrix}\right), \tag{15}$$

where $\boldsymbol{\mu}^R = (\mu_1^{RR}, \mu_1^{RA}, \mu_1^{RD}, \mu_J^{RR}, \mu_J^{RA}, \mu_J^{RD})'$ with

$$\begin{aligned}
\mu_1^{RR} &= -\sqrt{\frac{\left(n_{10}+n_{11}\right)n_{12}}{n_1}}\beta_1,\\
\mu_1^{RA} &= \frac{-n_{12}\left(2n_{10}+n_{11}\right)\beta_1}{\sqrt{n_1\left[n_{10}\left(n_{11}+4n_{12}\right)+n_{11}n_{12}\right]}},\\
\mu_1^{RD} &= \frac{-\sqrt{n_{10}}n_{12}\beta_1}{\sqrt{n_1\left(n_{11}+n_{12}\right)}},\\
\mu_J^{RR} &= -\sqrt{\frac{\left(N_0+N_1\right)N_2}{n}}\beta_1,\\
\mu_J^{RA} &= \frac{-N_2\left(2N_0+N_1\right)\beta_1}{\sqrt{n\left[N_0\left(N_1+4N_2\right)+N_1N_2\right]}},\\
\mu_J^{RD} &= \frac{-\sqrt{N_0}N_2\beta_1}{\sqrt{n\left(N_1+N_2\right)}},
\end{aligned} \tag{16}$$

(ii) when the true genetic model is additive,

$$\left(Z_1^R, Z_1^A, Z_1^D, Z_J^R, Z_J^A, Z_J^D\right)'\Big|_{H_1} \sim N_6\left(\boldsymbol{\mu}^A, \sigma^2\begin{pmatrix}\mathbf{V}_1 & \boldsymbol{\rho}\\ \boldsymbol{\rho}' & \mathbf{V}_J\end{pmatrix}\right), \tag{17}$$

where $\boldsymbol{\mu}^A = (\mu_1^{AR}, \mu_1^{AA}, \mu_1^{AD}, \mu_J^{AR}, \mu_J^{AA}, \mu_J^{AD})'$ with

$$\begin{aligned}
\mu_1^{AR} &= \frac{-\sqrt{n_{12}}\left(n_{11}+2n_{10}\right)\beta_1}{\sqrt{n_1\left(n_{10}+n_{11}\right)}},\\
\mu_1^{AA} &= -\sqrt{\frac{n_{10}\left(n_{11}+4n_{12}\right)+n_{11}n_{12}}{n_1}}\beta_1,\\
\mu_1^{AD} &= \frac{-\sqrt{n_{10}}\left(n_{11}+2n_{12}\right)\beta_1}{\sqrt{n_1\left(n_{11}+n_{12}\right)}},\\
\mu_J^{AR} &= \frac{-\sqrt{N_2}\left(N_1+2N_0\right)\beta_1}{\sqrt{n\left(N_0+N_1\right)}},
\end{aligned}$$

$$\mu_J^{AA} = -\sqrt{\frac{N_0\left(N_1+4N_2\right)+N_1N_2}{n}}\beta_1,$$
$$\mu_J^{AD} = \frac{-\sqrt{N_0}\left(N_1+2N_2\right)\beta_1}{\sqrt{n\left(N_1+N_2\right)}},$$
(18)

(iii) when the true genetic model is dominant,

$$\left(Z_1^R, Z_1^A, Z_1^D, Z_J^R, Z_J^A, Z_J^D\right)'\Big|_{H_1} \sim N_6\left(\boldsymbol{\mu}^D, \sigma^2\begin{pmatrix}\mathbf{V}_1 & \boldsymbol{\rho}\\ \boldsymbol{\rho}' & \mathbf{V}_J\end{pmatrix}\right),$$
(19)

where $\boldsymbol{\mu}^D = (\mu_1^{DR}, \mu_1^{DA}, \mu_1^{DD}, \mu_J^{DR}, \mu_J^{DA}, \mu_J^{DD})'$ with

$$\mu_1^{DR} = \frac{-\sqrt{n_{12}}n_{10}\beta_1}{\sqrt{n_1\left(n_{10}+n_{11}\right)}},$$
$$\mu_1^{DA} = \frac{-n_{10}\left(n_{11}+2n_{12}\right)\beta_1}{\sqrt{n_1\left[n_{10}\left(n_{11}+4n_{12}\right)+n_{11}n_{12}\right]}},$$
$$\mu_1^{DD} = -\sqrt{\frac{n_{10}\left(n_{11}+n_{12}\right)}{n_1}}\beta_1,$$
$$\mu_J^{DR} = \frac{-\sqrt{N_2}N_0\beta_1}{\sqrt{n\left(N_0+N_1\right)}},$$
$$\mu_J^{DA} = \frac{-N_0\left(N_1+2N_2\right)\beta_1}{\sqrt{n\left[N_0\left(N_1+4N_2\right)+N_1N_2\right]}},$$
$$\mu_J^{DD} = -\sqrt{\frac{N_0\left(N_1+N_2\right)}{n}}\beta_1.$$
(20)

We develop a method for simplifying the calculations of $\Pr_{H_0}(F_1^{\mathrm{MAX}} \le u_1)$ and $\Pr_{H_0}(F_J^{\mathrm{MAX}} \le u_J)$ and $\Pr_{H_0}(F_1^{\mathrm{MAX}} \le u_1, F_J^{\mathrm{MAX}} \le u_J)$. The details are included in Appendix B, and the calculations of $\Pr_{H_1}(F_1^{\mathrm{MAX}} \le u_1)$ and $\Pr_{H_1}(F_J^{\mathrm{MAX}} \le u_J)$ and $\Pr_{H_1}(F_1^{\mathrm{MAX}} \le u_1, F_J^{\mathrm{MAX}} \le u_J)$ are essentially similar.

3. Results

3.1. Power Comparison. We conduct simulation studies to evaluate the performance of the proposed method under three commonly used genetic models (recessive, additive, and dominant models). We mainly compare the power of two approaches; one is the proposed method in this paper, and the other is the joint analysis based on the F-test statistics F_1^A and F_J^A. For convenience, we refer to the proposed method as MAXFJ and AFJ for the other one. We choose the sample size $n = 2000$, and $m = 5 \times 10^5$. The proportion of subjects genotyped in Stage 1 has three levels $\pi = 0.3, 0.4, 0.5$. We set the genome-wide significance level as $\alpha = 0.05$ and that the significance level per SNP as $\alpha/m = 1 \times 10^{-7}$. In Stage 1, the P-value threshold for SNPs selected for followup is set to be 1×10^{-4} and 2×10^{-4}. We assume that the Hardy-Weinberg equilibrium holds in the general sample population, and then there are on average $n \times (1 - \mathrm{MAF})^2$, $2n \times \mathrm{MAF} \times (1 - \mathrm{MAF})$, and $n \times \mathrm{MAF}^2$ individuals with genotype gg, Gg, and GG, respectively, where the minor allele frequency is set to be 0.15, 0.30 and 0.45. To make the power comparison more distinctly, we specify different genetic effect parameters β_1 under three genetic models as follows: $\beta_1 = 0.5$ for the recessive model, $\beta_1 = 0.3$ for the additive model, and $\beta_1 = 0.4$ for the dominant model.

Table 1: Power comparison ($n = 2000$, $\gamma = 1 \times 10^{-4}$, $\alpha = 0.05$, and $m = 5 \times 10^5$).

π	MAF	REC		ADD		DOM	
		AFJ	MAXFJ	AFJ	MAXFJ	AFJ	MAXFJ
0.30	0.15	$7.5e-5$	0.005	0.426	0.365	0.610	0.618
	0.30	0.052	0.285	0.811	0.759	0.698	0.784
	0.45	0.487	0.785	0.893	0.854	0.449	0.647
0.40	0.15	$1.1e-4$	0.009	0.651	0.589	0.826	0.837
	0.30	0.086	0.470	0.945	0.922	0.887	0.938
	0.45	0.711	0.938	0.979	0.968	0.677	0.859
0.50	0.15	$1.0e-4$	0.010	0.802	0.751	0.933	0.941
	0.30	0.121	0.639	0.987	0.980	0.965	0.986
	0.45	0.856	0.987	0.997	0.995	0.826	0.953

Table 2: Power comparison ($n = 2000$, $\gamma = 2 \times 10^{-4}$, $\alpha = 0.05$, and $m = 5 \times 10^5$).

π	MAF	REC		ADD		DOM	
		AFJ	MAXFJ	AFJ	MAXFJ	AFJ	MAXFJ
0.30	0.15	$1.3e-4$	0.006	0.489	0.426	0.676	0.681
	0.30	0.066	0.340	0.852	0.806	0.754	0.828
	0.45	0.556	0.833	0.922	0.891	0.516	0.706
0.40	0.15	$1.2e-4$	0.011	0.709	0.651	0.866	0.876
	0.30	0.101	0.529	0.961	0.943	0.916	0.956
	0.45	0.765	0.957	0.987	0.979	0.732	0.892
0.50	0.15	$1.7e-4$	0.012	0.838	0.793	0.951	0.958
	0.30	0.133	0.683	0.992	0.987	0.975	0.991
	0.45	0.888	0.992	0.998	0.997	0.860	0.967

The power results are displayed in Tables 1 and 2 for $\gamma = 1 \times 10^{-4}$ and $\gamma = 2 \times 10^{-4}$, respectively. They indicate that MAXFJ is more efficiency robust than AFJ across various inheritance models. As expected, AFJ is more powerful than MAXFJ under the additive model. However, MAFJ performs much more powerful than AFJ when the true genetic model is recessive. For instance, in Table 2, with $\pi = 0.4$ and MAF = 0.3, the powers of AFJ and MAXFJ are 0.101 and 0.529, respectively. In summary, MAXFJ is substantially more powerful than AFJ in two-staged GWAS of quantitative traits, when the model for AFJ is misspecified.

3.2. An Illustration Example: Rheumatoid Arthritis. Rheumatoid arthritis (RA) is an autoimmune disease (resulting in a chronically systemic inflammatory disorder) which mainly attacks synovial joints. About 1% of the common adult population worldwide is affected by RA [16]. It has been

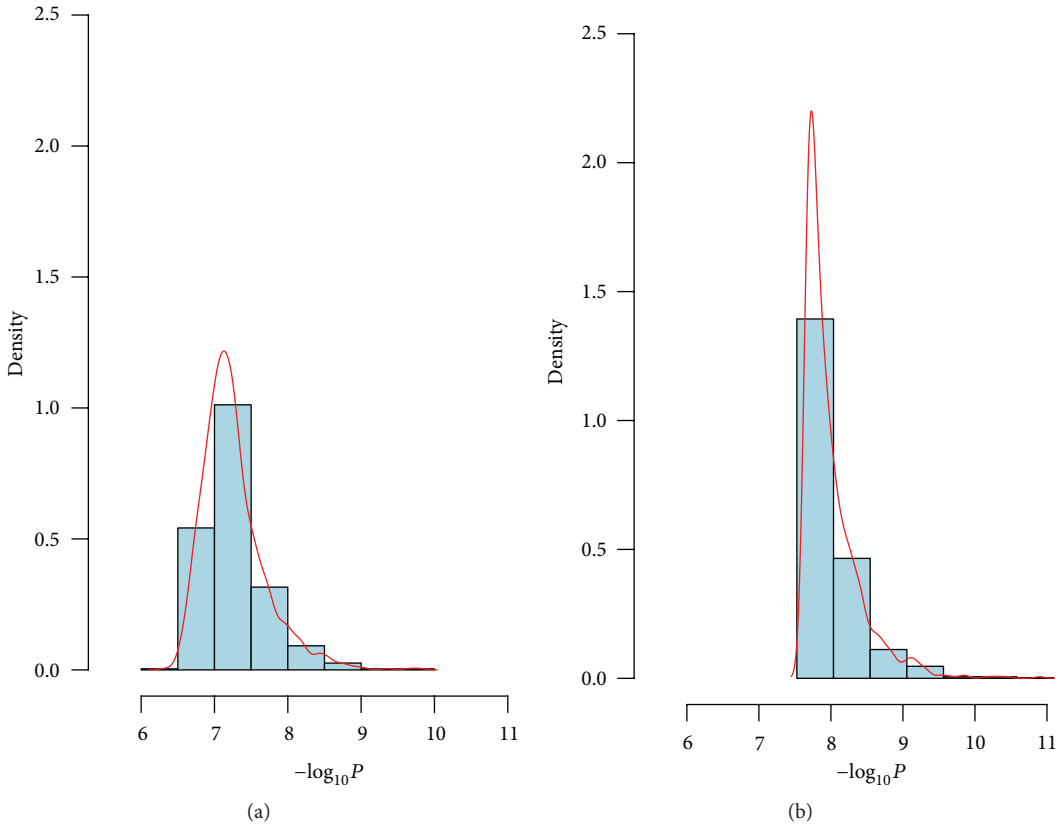

FIGURE 1: The histogram and density of $-\log_{10}P$ when $\pi = 0.3$ (the left subgraph corresponds to MAXFJ while the right one for AFJ).

pointed out that the genetic variants might play a major role in RA susceptibility [17]. Genetic Analysis Workshop 16 (GAW16) based on the North American Rheumatoid Arthritis Consortium (NARAC) is a GWAS testing association with RA using about 5×10^5 SNPs [18–20]. It included 868 individuals who were RA positive (cases) and also had continuous trait anticyclic citrullinated peptide (anti-CCP) measures and 1194 controls sampled from the New York Cancer Project (NYCP) without RA which had no anti-CCP measures. Huizinga et al. [21] pointed out that a greater anti-CCP would be linked to better prediction of increased risk developing RA. Chen et al. [22] showed that SNP rs2476601 located in PTPN22 had the most significant association with RA. Here, we only focus on SNP rs2476601 and apply two joint analysis methods (AFJ and MAXFJ) to evaluate its statistical significance. The minimum of anti-CCP among 868 cases was affected to each control, and a log transformation of anti-CCP was applied in the analysis. Then, we considered $\pi = 0.3, 0.4, 0.5$ three simulation circumstances. For $\pi = 0.3$, thirty percent of individuals were randomly sampled from all cases and controls and were used as the data from Stage 1, and the rest of individuals were treated as the data of Stage 2. The P-values of AFJ and MAXFJ were calculated, respectively. We repeated the above procedure 1,000 times and saved the corresponding P-values. A base-10 logarithm transformation and an opposite transformation were successively applied to these P-values, and the histogram and density of these transformed data were obtained (Figure 1). Similarly, we conducted the simulation and calculation for $\pi = 0.4$ and 0.5, and the corresponding histogram and density were presented in Figures 2 and 3. Examination of Figures 1–3 showed that the P-values of MAXFJ are more stable than those of AFJ and the estimated density curves of MAXFJ are more closer to the symmetrical normal distribution while the estimated density curves of AFJ are rather skewed, which indicated that MAXFJ possesses more robust performance when the real genetic models are unknown.

4. Discussion

We have developed a feasible two-stage design and the corresponding robust joint analysis approach for quantitative trait GWASs. The method is based on the F-statistics over three different genetic models. The denominator of the used F-statistic, which is constructed without assuming any genetic model, is different from the commonly used one. This adoption can reduce the computation intensity. Taking advantage of an ingenious design matrix, we successfully construct the common denominator of three F-test statistics for the joint analysis with combined raw data from both stages. The statistical significance (P-value) for the proposed joint analysis method can be calculated with the derived analytic expressions on the basis of the asymptotic distributions, which greatly reduce the complexity and computational intensity compared with the resampling-type permutation and bootstrap procedures. Our numerical results demonstrate

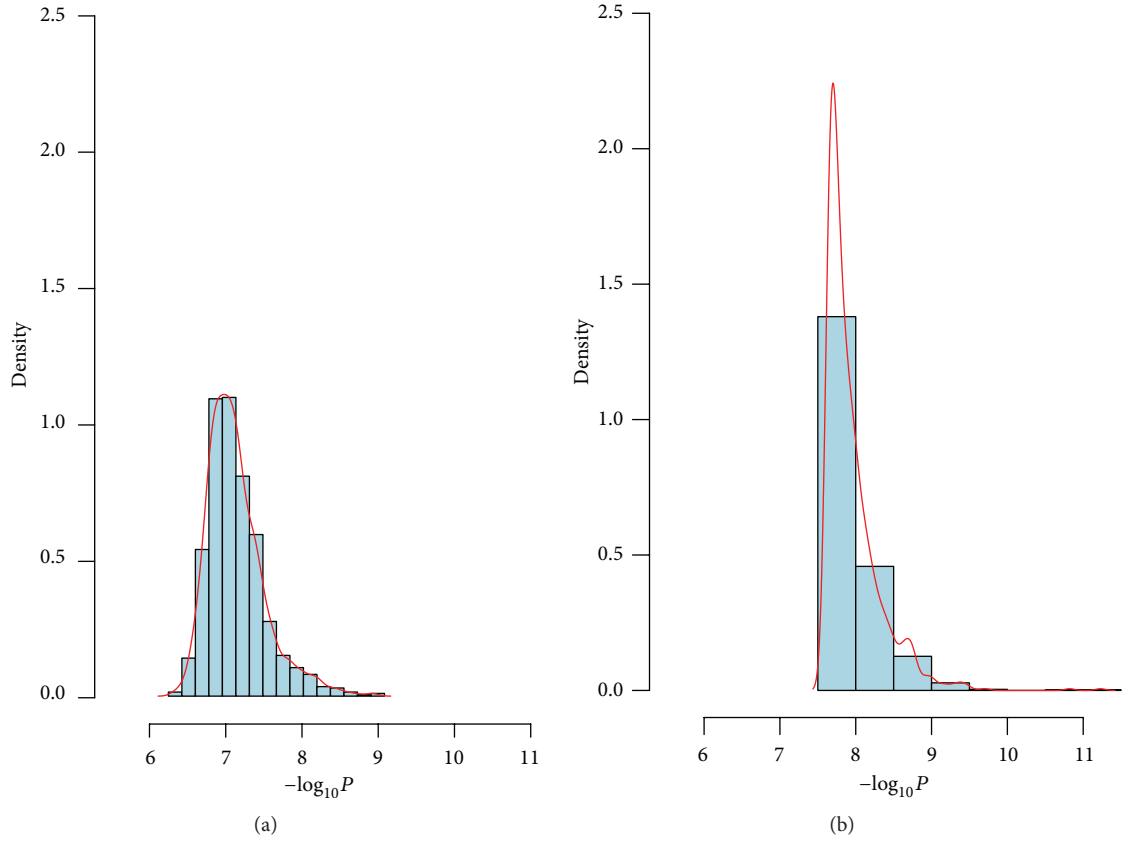

FIGURE 2: The histogram and density of $-\log_{10}P$ when $\pi = 0.4$ (the left subgraph corresponds to MAXFJ while the right one for AFJ).

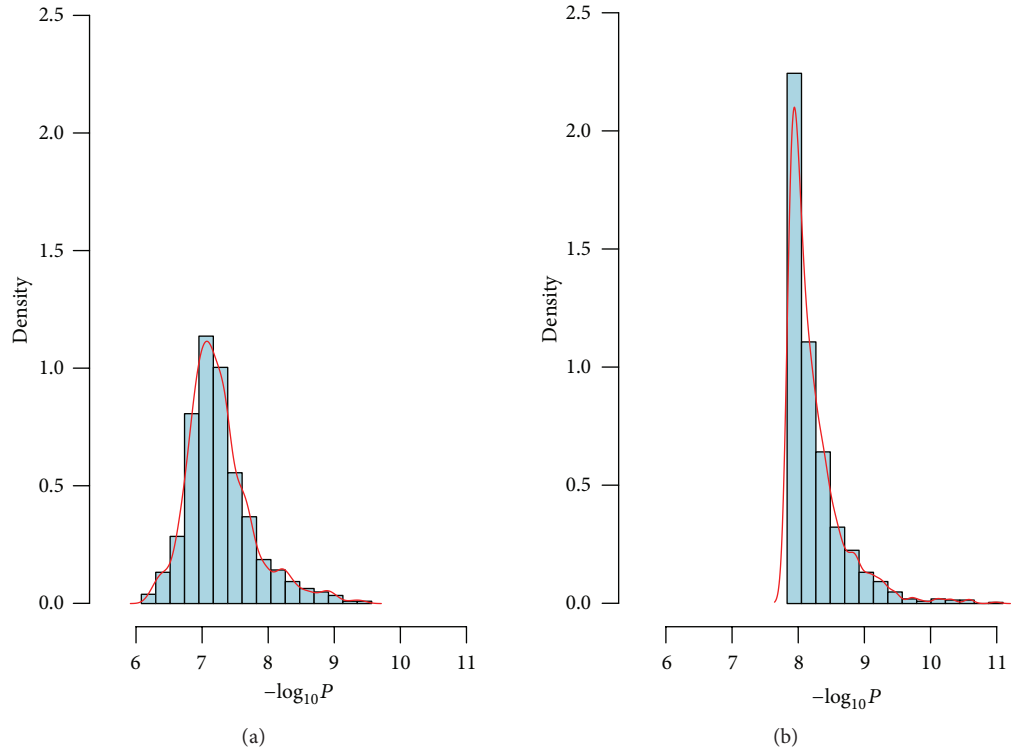

FIGURE 3: The histogram and density of $-\log_{10}P$ when $\pi = 0.5$ (the left subgraph corresponds to MAXFJ while the right one for AFJ).

that this novel approach has the greater efficiency robustness for genetic model uncertainty than the F-statistic-based joint analysis which assumes the additive genetic model.

In this work, we did not investigate the power of joint analysis based on other existing robust association methods for quantitative traits such as So and Sham's method. We find that it is very difficult to extend So and Sham's method (score test-based MAX3) to two-staged GWASs with quantitative outcomes, since it is almost impossible to derive the joint distribution of score tests from two stages.

For simplicity, here we do not take into account the effects of covariates in the considered two-stage design. However, in real application, the proposed method can be easily applied to the situation including one or more covariates as shown by the original MAXF by Li et al. [10]. It is important to stress that we combine the raw data from two stages to construct the joint statistic, unlike the joint analysis for binary traits using the weighted sum of two statistics in Stages 1 and 2 [12]. Furthermore, one basic assumption in this paper is that the effect sizes of genetic variants between two stages are identical (i.e., no heterogeneity exists), which is the natural and reasonable precondition for the data fusion strategy. In addition, the population-based genetic association studies may be affected by the population stratification, and this needs future research to examine it.

Appendix

A. The Derivation of the Asymptotic Properties of F_J^R, F_J^A, F_J^D

Consider the linear model for the combined raw data from Stage 1 and Stage 2 as follows:

$$\mathbf{Y} = \mathbf{W}\boldsymbol{\vartheta} + \boldsymbol{\varepsilon}, \tag{A.1}$$

where $\boldsymbol{\vartheta} = (\beta_0, \zeta_1, \zeta_2, \beta_0^*, \zeta_1^*, \zeta_2^*)'$ and $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$.

Denote

$$\mathbf{C}_R = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{C}_D = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{C}_A = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 2 & -1 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{C}_0 = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{A.2}$$

Based on the design matrix

$$\mathbf{W} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{O}_{n_1\times 3} \\ \mathbf{O}_{n_2\times 3} & \mathbf{X}_2 \end{pmatrix} \tag{A.3}$$

for the expanded full model above, we can get the ordinary least square estimator of $\boldsymbol{\vartheta}$ by $\widehat{\boldsymbol{\vartheta}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Y}$, and the residual sum of square is given by $\mathrm{RSS}_J = \mathbf{Y}'[\mathbf{I}_n - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}']\mathbf{Y}$. Furthermore, we denote the residual sum of squares under the following constraints: $\mathbf{C}_R\boldsymbol{\vartheta} = \mathbf{0}_4$, $\mathbf{C}_D\boldsymbol{\vartheta} = \mathbf{0}_4$, $\mathbf{C}_A\boldsymbol{\vartheta} = \mathbf{0}_4$, and $\mathbf{C}_0\boldsymbol{\vartheta} = \mathbf{0}_5$, by RSS_R, RSS_D, RSS_A, and RSS_0, respectively. After some algebras, we can obtain

$$F_J^R = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_R}{\mathrm{RSS}_J/(n-6)} = \frac{(\mathrm{RSS}_0 - \mathrm{RSS}_J) - (\mathrm{RSS}_R - \mathrm{RSS}_J)}{\mathrm{RSS}_J/(n-6)},$$

$$F_J^A = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_A}{\mathrm{RSS}_J/(n-6)} = \frac{(\mathrm{RSS}_0 - \mathrm{RSS}_J) - (\mathrm{RSS}_A - \mathrm{RSS}_J)}{\mathrm{RSS}_J/(n-6)},$$

$$F_J^D = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_D}{\mathrm{RSS}_J/(n-6)} = \frac{(\mathrm{RSS}_0 - \mathrm{RSS}_J) - (\mathrm{RSS}_D - \mathrm{RSS}_J)}{\mathrm{RSS}_J/(n-6)}. \tag{A.4}$$

On the one hand, according to

$$\mathbf{W}'\mathbf{W} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{O}_{3\times 3} \\ \mathbf{O}_{3\times 3} & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}, \tag{A.5}$$

it follows that

$$\left(\mathbf{W}'\mathbf{W}\right)^{-1} = \begin{pmatrix} \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} & \mathbf{O}_{3\times 3} \\ \mathbf{O}_{3\times 3} & \left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1} \end{pmatrix},$$

$$\mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}' = \begin{pmatrix} \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1' & \mathbf{O}_{n_1\times n_2} \\ \mathbf{O}_{n_2\times n_1} & \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2' \end{pmatrix}. \tag{A.6}$$

So, we can get that

$$\mathbf{Y}'\left[\mathbf{I}_n - \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'\right]\mathbf{Y} = \mathbf{Y}_1'\left[\mathbf{I}_{n_1} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{Y}_1 + \mathbf{Y}_2'\left[\mathbf{I}_{n_2} - \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\right]\mathbf{Y}_2. \tag{A.7}$$

That is, $\mathrm{RSS}_J = \mathrm{RSS}_1 + \mathrm{RSS}_2$. Furthermore, $\mathrm{RSS}_J/(n-6)$ is also the unbiased estimator of the variance of the residual σ^2 based on the independence and unbiasedness of RSS_1 and RSS_2.

On the other hand, $\mathrm{RSS}_0 - \mathrm{RSS}_J$ and $\mathrm{RSS}_R - \mathrm{RSS}_J$ are both independent of RSS_J, so $(Z_J^R)^2 = \mathrm{RSS}_0 - \mathrm{RSS}_R = (\mathrm{RSS}_0 - \mathrm{RSS}_J) - (\mathrm{RSS}_R - \mathrm{RSS}_J)$ is also independent of RSS_J, and $(\mathrm{RSS}_0 - \mathrm{RSS}_R)/\sigma^2 \sim \chi_1^2$, $\mathrm{RSS}_J/\sigma^2 \sim \chi_{n-6}^2$. Consequently, we have $F_J^R \sim F_{1,n-6}$. Similarly, we can get $F_J^A \sim F_{1,n-6}$ and $F_J^D \sim F_{1,n-6}$.

B. The Detailed Calculation of $\Pr_{H_0}(F_1^{\mathrm{MAX}} \le u_1)$ and $\Pr_{H_0}(F_J^{\mathrm{MAX}} \le u_J)$ and $\Pr_{H_0}(F_1^{\mathrm{MAX}} \le u_1, F_J^{\mathrm{MAX}} \le u_J)$

Denote $w_0 = \sqrt{(n_{10}+n_{11})n_{12}/(n_{10}(n_{11}+4n_{12})+n_{11}n_{12})}$ and $w_1 = \sqrt{n_{10}(n_{11}+n_{12})/(n_{10}(n_{11}+4n_{12})+n_{11}n_{12})}$. For a given $c > 0$,

$$\Pr_{H_0}\left(\left|\frac{Z_1^R}{\sigma}\right| \le c, \left|\frac{Z_1^A}{\sigma}\right| \le c, \left|\frac{Z_1^D}{\sigma}\right| \le c\right) = \oint_{\Omega_0} f(z_0, z_1; \boldsymbol{\Sigma}_0)\, dz_0 dz_1, \tag{B.1}$$

where $\Omega_0 = \{(z_0, z_1) : |z_0| \le c, |w_0 z_0 + w_1 z_1| \le c, |z_1| \le c\}$ and $f(z_0, z_1; \boldsymbol{\Sigma}_0)$ is the bivariate normal density function for $(Z_1^R/\sigma, Z_1^D/\sigma)'$ with mean $\mathbf{0}_2$ and variance-covariance matrix $\boldsymbol{\Sigma}_0 = \begin{pmatrix} 1 & \nu_{13} \\ \nu_{13} & 1 \end{pmatrix}$. Taking advantage of the symmetry of bivariate normal distribution, the above twofold integration can be only calculated at the right half space of Ω_0 and then multiplied by 2, which is

$$\begin{aligned}&\oint_{\Omega_0} f(z_0, z_1; \boldsymbol{\Sigma}_0)\, dz_0 dz_1 \\ &= 2\left[\int_0^{c(1-w_1)/w_0} dz_0 \int_{-c}^{c} f(z_0, z_1; \boldsymbol{\Sigma}_0)\, dz_1 \right. \\ &\quad \left. + \int_{c(1-w_1)/w_0}^{c} dz_0 \int_{-c}^{(c-w_0 z_0)/w_1} f(z_0, z_1; \boldsymbol{\Sigma}_0)\, dz_1\right].\end{aligned} \tag{B.2}$$

Based on the property of conditional distributions of the multivariate normal distribution, we have

$$f(z_0, z_1; \boldsymbol{\Sigma}_0) = \phi(z_0)\, f(z_1 \mid z_0; \nu_{13}), \tag{B.3}$$

where $\phi(z_0)$ is the probability density function of $N(0,1)$ and $f(z_1 \mid z_0; \nu_{13})$ is the density function of the conditional normal distribution $N(\nu_{13} z_0, 1-\nu_{13}^2)$. That is,

$$f(z_1 \mid z_0; \nu_{13}) = \frac{1}{\sqrt{1-\nu_{13}^2}} \phi\left(\frac{z_1 - \nu_{13} z_0}{\sqrt{1-\nu_{13}^2}}\right). \tag{B.4}$$

Then, it follows that

$$\begin{aligned}&\int_0^{c(1-w_1)/w_0} dz_0 \int_{-c}^{c} f(z_0, z_1; \boldsymbol{\Sigma}_0)\, dz_1 \\ &\quad = \int_0^{c(1-w_1)/w_0} \left[\Phi\left(\frac{c - \nu_{13} z_0}{\sqrt{1-\nu_{13}^2}}\right) - \Phi\left(\frac{-c - \nu_{13} z_0}{\sqrt{1-\nu_{13}^2}}\right)\right] \\ &\qquad \times \phi(z_0)\, dz_0, \\ &\int_{c(1-w_1)/w_0}^{c} dz_0 \int_{-c}^{(c-w_0 z_0)/w_1} f(z_0, z_1; \boldsymbol{\Sigma}_0)\, dz_1 \\ &\quad = \int_{c(1-w_1)/w_0}^{c} \left[\Phi\left(\frac{(c - w_0 z_0)/w_1 - \nu_{13} z_0}{\sqrt{1-\nu_{13}^2}}\right)\right. \\ &\qquad \left. - \Phi\left(\frac{-c - \nu_{13} z_0}{\sqrt{1-\nu_{13}^2}}\right)\right] \phi(z_0)\, dz_0.\end{aligned} \tag{B.5}$$

Thus,

$$\begin{aligned}&\oint_{\Omega_0} f(z_0, z_1; \boldsymbol{\Sigma}_0)\, dz_0 dz_1 \\ &= 2\left[\int_0^{c(1-w_1)/w_0} \Phi\left(\frac{c - \nu_{13} z_0}{\sqrt{1-\nu_{13}^2}}\right) \phi(z_0)\, dz_0 \right. \\ &\quad + \int_{c(1-w_1)/w_0}^{c} \Phi\left(\frac{(c - w_0 z_0)/w_2 - \nu_{13} z_0}{\sqrt{1-\nu_{13}^2}}\right) \\ &\qquad \times \phi(z_0)\, dz_0 \\ &\quad \left. - \int_0^{c} \Phi\left(\frac{-c - \nu_{13} z_0}{\sqrt{1-\nu_{13}^2}}\right) \phi(z_0)\, dz_0\right].\end{aligned} \tag{B.6}$$

Denote $w_0^* = \sqrt{(N_0+N_1)N_2/(N_0(N_1+4N_2)+N_1N_2)}$ and $w_1^* = \sqrt{N_0(N_1+N_2)/(N_0(N_1+4N_2)+N_1N_2)}$. For any given $c_1, c_2 > 0$,

$$\begin{aligned}&\Pr_{H_0}\left(\left|\frac{Z_1^R}{\sigma}\right| \le c_1, \left|\frac{Z_1^A}{\sigma}\right| \le c_1, \left|\frac{Z_1^D}{\sigma}\right| \le c_1, \right. \\ &\qquad \left. \left|\frac{Z_J^R}{\sigma}\right| \le c_2, \left|\frac{Z_J^A}{\sigma}\right| \le c_2, \left|\frac{Z_J^D}{\sigma}\right| \le c_2\right) \\ &\quad = \oint_{\Omega_1} f(z_0, z_1, z_2, z_3; \boldsymbol{\Sigma}_1)\, dz_0 dz_1 dz_2 dz_3,\end{aligned} \tag{B.7}$$

where

$$\Omega_1 = \{(z_0, z_1, z_2, z_3) : |z_0| \le c_1, |w_0 z_0 + w_1 z_1| \le c_1, \\ |z_1| \le c_1, |z_2| \le c_2, |w_0^* z_2 + w_1^* z_3| \le c_2, |z_3| \le c_2\} \tag{B.8}$$

and $f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)$ is the multivariate normal density function for $(Z_1^R/\sigma, Z_1^D/\sigma, Z_J^R/\sigma, Z_J^D/\sigma)'$ with mean $\mathbf{0}_4$ and variance-covariance matrix

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 1 & \nu_{13} & \rho_{11} & \rho_{13} \\ \nu_{13} & 1 & \rho_{31} & \rho_{33} \\ \rho_{11} & \rho_{31} & 1 & \nu_{13}^* \\ \rho_{13} & \rho_{33} & \nu_{13}^* & 1 \end{pmatrix}. \tag{B.9}$$

Note that $\oint_{\Omega_1} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1) dz_0 dz_1 dz_2 dz_3$ is the sum of nine integrations as follows:

$$L_1 = \int_{-c_1(1-w_1)/w_0}^{c_1(1-w_1)/w_0} dz_0 \int_{-c_1}^{c_1} dz_1 \int_{-c_2(1-w_1^*)/w_0^*}^{c_2(1-w_1^*)/w_0^*} dz_2 \int_{-c_2}^{c_2} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3,$$

$$L_2 = \int_{-c_1(1-w_1)/w_0}^{c_1(1-w_1)/w_0} dz_0 \int_{-c_1}^{c_1} dz_1 \int_{-c_2}^{-c_2(1-w_1^*)/w_0^*} dz_2 \int_{-(c_2+w_0^* z_2)/w_1^*}^{c_2} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3,$$

$$L_3 = \int_{-c_1(1-w_1)/w_0}^{c_1(1-w_1)/w_0} dz_0 \int_{-c_1}^{c_1} dz_1 \int_{c_2(1-w_1^*)/w_0^*}^{c_2} dz_2 \int_{-c_2}^{(c_2-w_0^* z_2)/w_1^*} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3,$$

$$L_4 = \int_{c_1(1-w_1)/w_0}^{c_1} dz_0 \int_{-c_1}^{(c_1-w_0 z_0)/w_1} dz_1 \int_{-c_2(1-w_1^*)/w_0^*}^{c_2(1-w_1^*)/w_0^*} dz_2 \int_{-c_2}^{c_2} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3,$$

$$L_5 = \int_{c_1(1-w_1)/w_0}^{c_1} dz_0 \int_{-c_1}^{(c_1-w_0 z_0)/w_1} dz_1 \int_{-c_2}^{-c_2(1-w_1^*)/w_0^*} dz_2 \int_{-(c_2+w_0^* z_2)/w_1^*}^{c_2} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3,$$

$$L_6 = \int_{c_1(1-w_1)/w_0}^{c_1} dz_0 \int_{-c_1}^{(c_1-w_0 z_0)/w_1} dz_1 \int_{c_2(1-w_1^*)/w_0^*}^{c_2} dz_2 \int_{-c_2}^{(c_2-w_0^* z_2)/w_1^*} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3,$$

$$L_7 = \int_{-c_1}^{-c_1(1-w_1)/w_0} dz_0 \int_{-(c_1+w_0 z_0)/w_1}^{c_1} dz_1 \int_{-c_2(1-w_1^*)/w_0^*}^{c_2(1-w_1^*)/w_0^*} dz_2 \int_{-c_2}^{c_2} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3,$$

$$L_8 = \int_{-c_1}^{-c_1(1-w_1)/w_0} dz_0 \int_{-(c_1+w_0 z_0)/w_1}^{c_1} dz_1 \int_{-c_2}^{-c_2(1-w_1^*)/w_0^*} dz_2 \int_{-(c_2+w_0^* z_2)/w_1^*}^{c_2} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3,$$

$$L_9 = \int_{-c_1}^{-c_1(1-w_1)/w_0} dz_0 \int_{-(c_1+w_0 z_0)/w_1}^{c_1} dz_1 \int_{c_2(1-w_1^*)/w_0^*}^{c_2} dz_2 \int_{-c_2}^{(c_2-w_0^* z_2)/w_1^*} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1)\, dz_3. \tag{B.10}$$

Moreover, we can obtain that $L_2 = L_3$, $L_4 = L_7$, $L_5 = L_9$, and $L_6 = L_8$ based on the symmetry of the integration domain for (z_0, z_1) and (z_2, z_3), respectively.

We have

$$\begin{aligned} f(z_0, z_1, z_2, z_3; \mathbf{\Sigma}_1) &= \phi(z_0) f(z_1 \mid z_0; \nu_{13}) \\ &\quad \times f(z_2 \mid z_0, z_1; \nu_{13}, \rho_{11}, \rho_{31}) \\ &\quad \times f(z_3 \mid z_0, z_1, z_2; \nu_{13}, \nu_{13}^*, \rho_{11}, \rho_{13}, \rho_{31}, \rho_{33}), \end{aligned} \tag{B.11}$$

where $f(z_2 \mid z_0, z_1; \nu_{13}, \rho_{11}, \rho_{31})$ is the conditional normal density function as

$$\begin{aligned} &f(z_2 \mid z_0, z_1; \nu_{13}, \rho_{11}, \rho_{31}) \\ &= \frac{1}{\sqrt{1 - \left(\left(\rho_{11}^2 - 2\rho_{11}\rho_{31}\nu_{13} + \rho_{31}^2\right) / \left(1 - \nu_{13}^2\right)\right)}} \\ &\quad \times \phi\Bigg[\Big(\Big(z_2 - \big[z_0 \left(\rho_{11} - \nu_{13}\rho_{31}\right) \\ &\qquad + z_1 \left(\rho_{31} - \rho_{11}\nu_{13}\right) \big] \Big) / \left(1 - \nu_{13}^2\right) \Big) \\ &\quad \times \Big(\sqrt{1 - \left(\left(\rho_{11}^2 + 2\rho_{11}\rho_{31}\nu_{13} + \rho_{31}^2\right) / \left(1 - \nu_{13}^2\right)\right)} \Big)^{-1} \Bigg], \end{aligned} \tag{B.12}$$

and $f(z_3 \mid z_0, z_1, z_2; \nu_{13}, \nu_{13}^*, \rho_{11}, \rho_{13}, \rho_{31}, \rho_{33})$ is the conditional normal density function given by

$$f(z_3 \mid z_0, z_1, z_2; \nu_{13}, \nu_{13}^*, \rho_{11}, \rho_{13}, \rho_{31}, \rho_{33}) = \frac{1}{\sqrt{1-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(\rho_{13}, \rho_{33}, \nu_{13}^*)'}} \times \phi\left(\frac{z_3 - (\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'}{\sqrt{1-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(\rho_{13}, \rho_{33}, \nu_{13}^*)'}}\right) \tag{B.13}$$

with $\boldsymbol{\Gamma}$ the submatrix of $\boldsymbol{\Sigma}_1$ formed by first three rows and three columns.
Denote $(\sigma^*)^2 = (\rho_{13}, \rho_{33}, \nu_{13}^*)\boldsymbol{\Gamma}^{-1}(\rho_{13}, \rho_{33}, \nu_{13}^*)'$. Then, we have

$$\begin{aligned}
L_1 &= \int_{-c_1(1-w_1)/w_0}^{c_1(1-w_1)/w_0} \phi(z_0)\,dz_0 \int_{-c_1}^{c_1} \frac{1}{\sqrt{1-\nu_{13}^2}}\phi\left(\frac{z_1-\nu_{13}z_0}{\sqrt{1-\nu_{13}^2}}\right)dz_1 \int_{-c_2(1-w_1^*)/w_0^*}^{c_2(1-w_1^*)/w_0^*} \left[\Phi\left(\left(c_2-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right. \\
&\quad \left. -\Phi\left(\left(-c_2-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right]dz_2, \\
L_2 &= \int_{-c_1(1-w_1)/w_0}^{c_1(1-w_1)/w_0} \phi(z_0)\,dz_0 \int_{-c_1}^{c_1} \frac{1}{\sqrt{1-\nu_{13}^2}}\phi\left(\frac{z_1-\nu_{13}z_0}{\sqrt{1-\nu_{13}^2}}\right)dz_1 \int_{-c_2}^{-c_2(1-w_1^*)/w_0^*} \left[\Phi\left(\left(c_2-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right. \\
&\quad \left. -\Phi\left(\left(-\frac{c_2+w_0^*z_2}{w_1^*}-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right]dz_2, \\
L_4 &= \int_{c_1(1-w_1)/w_0}^{c_1} \phi(z_0)\,dz_0 \int_{-c_1}^{(c_1-w_0z_0)/w_1} \frac{1}{\sqrt{1-\nu_{13}^2}}\phi\left(\frac{z_1-\nu_{13}z_0}{\sqrt{1-\nu_{13}^2}}\right)dz_1 \int_{-c_2(1-w_1^*)/w_0^*}^{c_2(1-w_1^*)/w_0^*} \left[\Phi\left(\left(c_2-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right. \\
&\quad \left. -\Phi\left(\left(-c_2-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right]dz_2, \\
L_5 &= \int_{c_1(1-w_1)/w_0}^{c_1} \phi(z_0)\,dz_0 \int_{-c_1}^{(c_1-w_0z_0)/w_1} \frac{1}{\sqrt{1-\nu_{13}^2}}\phi\left(\frac{z_1-\nu_{13}z_0}{\sqrt{1-\nu_{13}^2}}\right)dz_1 \int_{-c_2}^{-c_2(1-w_1^*)/w_0^*} \left[\Phi\left(\left(c_2-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right. \\
&\quad \left. -\Phi\left(\left(-\frac{c_2+w_0^*z_2}{w_1^*}-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right]dz_2, \\
L_6 &= \int_{c_1(1-w_1)/w_0}^{c_1} \phi(z_0)\,dz_0 \int_{-c_1}^{(c_1-w_0z_0)/w_1} \frac{1}{\sqrt{1-\nu_{13}^2}}\phi\left(\frac{z_1-\nu_{13}z_0}{\sqrt{1-\nu_{13}^2}}\right)dz_1 \int_{c_2(1-w_1^*)/w_0^*}^{c_2} \left[\Phi\left(\left(\frac{c_2-w_0^*z_2}{w_1^*}-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right. \\
&\quad \left. -\Phi\left(\left(-c_2-(\rho_{13}, \rho_{33}, \nu_{13}^*)\,\boldsymbol{\Gamma}^{-1}(z_0, z_1, z_2)'\right)\left(\sqrt{1-(\sigma^*)^2}\right)^{-1}\right)\right]dz_2.
\end{aligned} \tag{B.14}$$

Finally,

$$\oint_{\Omega_1} f(z_0, z_1, z_2, z_3; \boldsymbol{\Sigma}_1)\,dz_0dz_1dz_2dz_3 = L_1 + 2(L_2 + L_4 + L_5 + L_6). \tag{B.15}$$

Acknowledgments

The work was partially supported by the National Natural Science Foundation of China (no. 11225103; 11161054; 11171293; 81072386); the Key Fund of Yunnan Province (no. 2010CC003); The Youth Program of Applied Basic Research

Programs of Yunnan Province (2013FD001); the Foundation of Yunnan University (no. 2012CG018).

References

[1] B. Freidlin, G. Zheng, Z. Li, and J. L. Gastwirth, "Trend tests for case-control studies of genetic markers: power, sample size and robustness," *Human Heredity*, vol. 53, no. 3, pp. 146–152, 2002.

[2] J. D. Storey and R. Tibshirani, "Statistical significance for genomewide studies," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 16, pp. 9440–9445, 2003.

[3] K. Song and R. C. Elston, "A powerful method of combining measures of association and Hardy-Weinberg disequilibrium for fine-mapping in case-control studies," *Statistics in Medicine*, vol. 25, no. 1, pp. 105–126, 2006.

[4] G. Zheng and J. L. Gastwirth, "On estimation of the variance in Cochran-Armitage trend tests for genetic association using case-control studies," *Statistics in Medicine*, vol. 25, no. 18, pp. 3150–3159, 2006.

[5] Q. Li, K. Yu, Z. Li, and G. Zheng, "MAX-rank: a simple and robust genome-wide scan for case-control association studies," *Human Genetics*, vol. 123, no. 6, pp. 617–623, 2008.

[6] Q. Li, G. Zheng, X. Liang, and K. Yu, "Robust tests for single-marker analysis in case-control genetic association studies," *Annals of Human Genetics*, vol. 73, no. 2, pp. 245–252, 2009.

[7] Y. Zang and W. K. Fung, "Robust tests for matched case-control genetic association studies," *BMC Genetics*, vol. 11, article 91, 2010.

[8] Y. Zang, W. K. Fung, and G. Zheng, "Simple algorithms to calculate asymptotic null distributions of robust tests in case-control genetic association studies in R," *Journal of Statistical Software*, vol. 33, no. 8, pp. 1–24, 2010.

[9] H. C. So and P. C. Sham, "Robust association tests under different genetic models, allowing for binary or quantitative traits and covariates," *Behavior Genetics*, vol. 41, no. 5, pp. 768–775, 2011.

[10] Q. Li, W. Xiong, J. B. Chen et al., "A robust test for quantitative trait analysis with model uncertainty in genetic association studies," *Statistics and Its Interface*. In press.

[11] J. M. Satagopan, E. S. Venkatraman, and C. B. Begg, "Two-stage designs for gene-disease association studies with sample size constraints," *Biometrics*, vol. 60, no. 3, pp. 589–597, 2004.

[12] A. D. Skol, L. J. Scott, G. R. Abecasis, and M. Boehnke, "Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies," *Nature Genetics*, vol. 38, no. 2, pp. 209–213, 2006.

[13] K. Yu, N. Chatterjee, W. Wheeler et al., "Flexible design for following up positive findings," *American Journal of Human Genetics*, vol. 81, no. 3, pp. 540–551, 2007.

[14] R. Sladek, G. Rocheleau, J. Rung et al., "A genome-wide association study identifies novel risk loci for type 2 diabetes," *Nature*, vol. 445, no. 7130, pp. 881–885, 2007.

[15] D. Pan, Q. Li, N. Jiang, A. Liu, and K. Yu, "Robust joint analysis allowing for model uncertainty in two-stage genetic association studies," *BMC Bioinformatics*, vol. 12, article 9, 2011.

[16] A. J. Silman and J. E. Pearson, "Epidemiology and genetics of rheumatoid arthritis," *Arthritis Research & Therapy*, vol. 4, supplement 3, pp. S265–S272, 2002.

[17] A. J. MacGregor, H. Snieder, A. S. Rigby et al., "Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins," *Arthritis & Rheumatism*, vol. 43, no. 1, pp. 30–37, 2000.

[18] C. I. Amos, W. V. Chen, M. F. Seldin et al., "Data for Genetic Analysis Workshop 16 Problem 1, association analysis of rheumatoid arthritis data," *BMC Proceedings*, vol. 3, supplement 7, article S2, 2009.

[19] F. Xia, J. Y. Zhou, and W. K. Fung, "A powerful approach for association analysis incorporating imprinting effects," *Bioinformatics*, vol. 27, no. 18, pp. 2571–2577, 2011.

[20] G. Zheng, C. O. Wu, M. Kwak, W. Jiang, J. Joo, and J. A. C. Lima, "Joint analysis of binary and quantitative traits with data sharing and outcome-dependent sampling," *Genetic Epidemiology*, vol. 36, no. 3, pp. 263–273, 2012.

[21] T. W. J. Huizinga, C. I. Amos, A. H. M. van der Helm-Van Mil et al., "Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins," *Arthritis & Rheumatism*, vol. 52, no. 11, pp. 3433–3438, 2005.

[22] L. Chen, M. Zhong, W. V. Chen, C. I. Amos, and R. Fan, "A genome-wide association scan for rheumatoid arthritis data by Hotellings T^2 tests," *BMC Proceedings*, vol. 3, supplement 7, article S6, 2009.