Genome analysis

Controlling the joint local false discovery rate is more powerful than meta-analysis methods in joint analysis of summary statistics from multiple genome-wide association studies

Wei Jiang and Weichuan Yu*

Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on August 4, 2016; revised on October 6, 2016; editorial decision on October 28, 2016; accepted on October 30, 2016

Abstract

Motivation: In genome-wide association studies (GWASs) of common diseases/traits, we often analyze multiple GWASs with the same phenotype together to discover associated genetic variants with higher power. Since it is difficult to access data with detailed individual measurements, summary-statistics-based meta-analysis methods have become popular to jointly analyze datasets from multiple GWASs.

Results: In this paper, we propose a novel summary-statistics-based joint analysis method based on controlling the joint local false discovery rate (Jlfdr). We prove that our method is the most powerful summary-statistics-based joint analysis method when controlling the false discovery rate at a certain level. In particular, the Jlfdr-based method achieves higher power than commonly used meta-analysis methods when analyzing heterogeneous datasets from multiple GWASs. Simulation experiments demonstrate the superior power of our method over meta-analysis methods. Also, our method discovers more associations than meta-analysis methods from empirical datasets of four phenotypes.


Contact: eeyu@ust.hk

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Understanding genetic mechanisms of common diseases and traits is important in biological and medical research. The goal of genome-wide association studies (GWASs) is to associate the variation in single nucleotide polymorphisms (SNPs) with variation in common diseases/traits (Altshuler et al., 2008). Due to decreasing genotyping costs (Perkel, 2008), constantly emerging successful stories (Klein et al., 2003; Kraft and Haiman, 2010) and efforts of the GWAS consortia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Wellcome Trust Case Control Consortium, 2007), more and more GWASs have been conducted for common phenotypes (Welter et al., 2014).

Analyses of GWAS results show that the identified associations can only explain a small part of the additive genetic variances. This is referred to as the ‘missing heritability’ problem (Manolio et al., 2009). The hints of hidden heritability (Gibson, 2010; Yang et al., 2010) and the estimated distribution of common SNPs’ effect sizes (Park et al., 2010) suggest that common diseases/traits are influenced by thousands of SNPs with small effects. To discover these genetic variants with small effects, we need to improve studies’
Our major contribution in this paper is that we propose a novel method of joint analysis and summary-statistics-based joint analysis. We will use a z-value scheme to detect associations between SNPs and the phenotype, i.e. the test statistics follow a standard normal distribution under no hypothesis. We use \( \hat{\mu} \) to denote the observed effect size in study \( j \). The asymptotically standard error of \( \hat{\mu} \) is \( \sigma(\hat{\mu}) \). Correspondingly, the test statistic in study \( j \) is \( z(\hat{\mu}) = \frac{\hat{\mu}}{\sigma(\hat{\mu})} \). The underlying expected effect size is \( \mu \). The expected effect size of the same SNP may vary in different studies due to heterogeneity. The test statistic \( Z(\mu) \) (uppercase letter indicates a random variable) follows an \( N(\mu, \sigma^2(\mu), 1) \) distribution. We use \( z \) to represent the vector of test statistics in all studies, i.e. \( z = (z(1), z(2))' \). Similarly, we use \( \mu \) to represent the vector of expected effect sizes in all studies, i.e. \( \mu = (\mu(1), \mu(2))' \).

We further assume \( m_0 \) SNPs have no association with the phenotype and \( m_1 \) SNPs have associations. Thus, the null proportion reads \( \pi_0 = m_0 / m (0 \leq \pi_0 \leq 1) \). We use \( H_0 \) and \( H_1 \) to denote the null hypothesis and the alternative hypothesis, respectively.

In the joint analysis of summary statistics from multiple GWASs, we assume that \( R \) of the \( m \) hypotheses are rejected. There are \( V \) false positives and \( S \) true positives (i.e. \( V + S = R \)). Table 1 summarizes the numbers of hypotheses in the different categories.

When testing multiple hypotheses, it is very easy to have false positives by random chance. This problem is known as the ‘multiplicity’ problem. Many criteria are proposed to address the multiplicity problem. We present an incomplete list of these criteria in Table 2(a). Let’s define the false discovery proportion (Fdp) as \( \frac{V}{R(1 + V)} \) with ‘v’ denoting the maximum operation. Fdp is an unknown quantity in real cases. The classical false discovery rate (FDR) is the expectation of the Fdp. Controlling the FDR is more powerful than controlling the family-wise error rate (FWER). The Bayesian false discovery rate (Fdr) is the expected value of the Fdp given \( R > 0 \). Compared to FDR, Fdr is conditional on \( R > 0 \) since we are only interested in controlling false positives when \( R > 0 \). We adopt Fdr in this paper as the criterion to avoid a plethora of false positives.

In addition to controlling false positives, we also need a criterion to measure the amount of true positives when evaluating a rejection region. A direct concept is power. The classical definition of power is a function of a given effect size as shown in the first row of Table 2(b). Since effect sizes of associated SNPs are different and unobserved, the actual power values are unknown. The Bayesian power removes the dependence of power on effect size by taking the expectation of the empirical power, which is defined as \( \frac{S}{m_1} \) (\( m_1 > 0 \)). We list the definitions of the power and the Bayesian power. Jointly analyzing datasets from multiple GWASs on the same disease in the same population provides an opportunity to improve the power.

There are two kinds of joint analysis methods: individual-level joint analysis and summary-statistics-based joint analysis. Individual-level joint analysis uses individual-level genotype data from all studies. One such example is mega-analysis (Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, 2013), which pools all data together. Summary-statistics-based joint analysis only uses summary statistics from different studies. Since individual-level genotype data is difficult to access, summary-statistics-based analysis is widely used in joint analysis. The most commonly used method of summary-statistics-based joint analysis is meta-analysis (Evangelou and Ioannidis, 2013), which derives a new statistic for each SNP using summary statistics from multiple studies.

Our focus in this paper is to study summary-statistics-based joint analysis methods. More specifically, we like to study which joint analysis method provides the highest power for a given false discovery rate level. Figure 1 illustrates our motivation.

Our major contribution in this paper is that we propose a novel summary-statistics-based joint analysis method based on controlling the joint local false discovery rate (Jlfdr). The Jlfdr generalizes the concept of the local false discovery rate (Efron, 2005) from the analysis of single study to the joint analysis of multiple studies. We will give implementation details of the Jlfdr-based method under the Gaussian mixture model. We will also discuss the relationship between the Jlfdr-based method and meta-analysis methods.

### 2 Methods

#### 2.1 Notations and criteria

Our method deals with a multiple GWAS setting. For simplicity, we illustrate the concepts with a two-GWAS setting. We use parenthesis superscript ‘(j)’ to denote the study index. For example, the sample sizes in study 1 and 2 are \( n^{(1)} \) and \( n^{(2)} \), respectively. We use subscript \( i \) (\( i = 1, \ldots, n \)) in study \( j \) to represent the study index. For example, the \( j \)-th study index.

#### 2.2 Methods

<table>
<thead>
<tr>
<th>( H_0 ) is true</th>
<th>( H_0 ) is false</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 ) is rejected</td>
<td>( V )</td>
<td>( S )</td>
</tr>
<tr>
<td>( H_0 ) is not rejected</td>
<td>( U )</td>
<td>( T )</td>
</tr>
</tbody>
</table>

The letter in each cell denotes the count of the hypotheses in each category.
power in Table 2(b). In this paper, we use the Bayesian power as the criterion to measure the amount of true positives.

Both Fdr and Bayesian power are functions of the rejection region $R$. For two different rejection regions with the same Fdr level, we prefer the region with higher Bayesian power because it can find more true associations without increasing the proportion of false positives in the findings. Thus, we propose a joint analysis method determining the optimal rejection region when controlling the Fdr at a certain threshold $q$, i.e.

$$\max_{R} \eta(R) \quad \text{s.t.} \quad \text{Fdr}(R) \leq q. \quad (1)$$

Here $\eta(R)$ denotes the Bayesian power. Actually, when controlling the Fdr at the same threshold, meta-analysis methods can also be regarded as the solutions to the above optimization problem with further constraint about the form of $\eta(R)$, i.e.

$$\max_{R_c} \eta(R_c) \quad \text{s.t.} \quad \text{Fdr}(R_c) \leq q \quad (2)$$

$R_c = \{z \mid g(z, z) \geq C\}$

Here $z = (\sqrt{u_1}, \sqrt{u_2})^T$ and $g$ is a function which has different forms in different meta-analysis methods. We will give the explicit forms of the function $g$ in meta-analysis methods in the Supplementary Note. Also, we will discuss the relationship between our proposed method and meta-analysis methods in detail in Section 2.4. In the next subsection, we will present the solution to the optimization problem in Eq. (1).

### 2.2 Jlfdr and optimal rejection region

To derive the solution to the optimization problem in Eq. (1), we need to introduce the concept of joint local false discovery rate (Jlfdr) first. Jlfdr is a simple extension of the local false discovery rate (Efron, 2005) from the analysis of single study to the joint analysis of multiple studies. It reads as

$$\text{Jlfdr}(z) = P(\mathcal{H}_0 \mid z), \quad (3)$$

which is the posterior probability of a null hypothesis, given the observed summary statistic vector $z$.

The relationship between Jlfdr and Fdr is (see the Supplementary Note for details)

$$\text{Fdr}(R) = E(\text{Jlfdr}(z) \mid z \in R). \quad (4)$$

In other words, Fdr is the expectation of Jlfdr, given that the test statistic vector is in the rejection region $R$.

Let us define a rejection region $R_C = \{z \mid \text{Jlfdr}(z) \leq t(q)\}$, where $t(q)$ is a threshold such that $\text{Fdr}(R_C) = q$. We have the following theorem:

**Theorem 1** For any rejection region $R$ with $\text{Fdr}(R) \leq q$, we have $\eta(R) \leq \eta(R_C)$.

We show the proof of Theorem 1 in the Supplementary Note. Theorem 1 shows that $R_C$ is the most powerful rejection region when controlling Fdr at $q$. This gives us a clue that we can improve the power of summary-statistics-based joint analysis by controlling the Jlfdr. In the next section, we shall provide details of the implementation of the Jlfdr-based method under the Gaussian mixture model.

### 2.3 Implementation of Jlfdr-based method under the Gaussian mixture model

We assume the distribution of $Z$ is a $(K + 1)$-component Gaussian mixture model:

$$Z \sim \pi_0 N(0, I) + \sum_{k=1}^{K} \pi_k N(0, I + \Sigma_k), \quad (5)$$

where $\sum_{k=1}^{K} \pi_k = 1 - \pi_0$. We provide the justification of making this assumption in the Supplementary Note. There are some unknown parameters $\pi_1 = (\pi_1, \ldots, \pi_k)^T$ and $\Sigma = (\Sigma_1, \ldots, \Sigma_k)$ in the above mixture model. Dempster et al. (1977) proposed an EM-algorithm to estimate parameters with unobserved latent variables. With the observed vectors of summary statistics $z_i$ ($i = 1, \ldots, m$), we use the EM-algorithm to estimate the parameters $\pi_1$ and $\Sigma$ in the Gaussian mixture model (5). Please note that $\pi_0$ is always much larger than any entry of $\pi_1$ in the GWAS setting. Hence, a Dirichlet($\pi_0, 0^T$) prior is added for the proportions ($\pi_0, \pi_1^T$). This is the same penalty strategy proposed by Muralidharan (2010).

Denote the probability density function (pdf) of bivariate normal distribution $N_2(0, I)$ as $f_0(x_1, x_2)$ and the pdf of $N_2(0, I + \Sigma_k)$ as $f_k(x_1, x_2 | \Sigma_k)$. The Jlfdr reads

$$\text{Jlfdr}(z) = \frac{\pi_0 f_0(z_1, z_2)}{\pi_0 f_0(z_1, z_2) + \sum_{k=1}^{K} \pi_k f_k(z_1, z_2 | \Sigma_k)} \quad (6)$$

After calculating the Jlfdr, we approximate Fdr as

$$\text{Fdr}(R) = E(\text{Jlfdr}(z) \mid z \in R) \approx \frac{1}{|\{z \in R\}|} \sum_{z \in R} \text{Jlfdr}(z). \quad (7)$$

We determine the optimal rejection region $R_C$ by Jlfdr-thresholding, which determines the rejection region with Jlfdr smaller than the threshold $t(q)$. To determine the threshold $t(q)$, we sort the

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Mathematical Definitions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Different criteria for controlling false positives in multiple testing scenario</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False discovery rate (FDR)</td>
<td>$\text{FDR}(R) = E(V</td>
<td>V \leq 1)$</td>
</tr>
<tr>
<td>Bayesian false discovery rate (BFR)</td>
<td>$\text{Fdr}(R) = E(V/V \neq 1)$</td>
<td>Storey (2003)</td>
</tr>
<tr>
<td>(b) Different criteria for measuring the amount of true positives in multiple testing scenario</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>$\beta(R, \mu) = P(z \in R \mid H_1, \mu)$</td>
<td>Neyman and Pearson (1933)</td>
</tr>
<tr>
<td>Bayesian power</td>
<td>$\eta(R) = P(z \in R \mid H_1)$</td>
<td>Kruschke (2010)</td>
</tr>
</tbody>
</table>

Here $R$ is the rejection region in the analysis, $\mu$ denotes effect sizes, $V$ and $R$ as well as other notations are explained in Table 1. ‘$\vee$’ denotes the maximum operation.
calculated Jlfdr values of each SNP in an ascending order first. Denote the $a$-th Jlfdr value as Jlfdr$_a$. We can approximate the Fdr of the region R$_a = \{z : \text{Jlfdr}(z) \leq \text{Jlfdr}_a\}$ as

$$\text{Fdr}(R_a) \approx \frac{1}{a} \sum_{i=1}^{a} \text{Jlfdr}_i.$$  

(8)

We use $c$ to denote the largest $a$ such that Fdr(R$_a$) $\leq q$, namely

$$c = \max\{a : \text{Fdr}(R_a) \leq q\}.$$  

(9)

Then the Jlfdr threshold $t(q)$ is Jlfdr$_c$. We reject all SNPs with Jlfdr(z) $\leq t(q)$.

We present the detailed steps of the Jlfdr-based method in Algorithm 2.3.

**Algorithm 1. Jlfdr-based method for summary-statistics-based joint analysis**

**Inference using the EM-algorithm:**

- Initialize $\pi_1$ and $\Sigma$
- repeat
  - **E Step:**
    $$\pi'_1 = 1 - \sum_{k=1}^{K} \pi'_{1k}$$
    $$b_{il} = \frac{\pi'_{1k} f(z_i^{(1)}, z_i^{(2)} | \pi'_{1k}, \Sigma_{il})}{\sum_{k=1}^{K} \pi'_{1k} f(z_i^{(1)}, z_i^{(2)} | \pi'_{1k}, \Sigma_{il})},$$
    $$b_{il} = \frac{\pi'_{1k} f(z_i^{(1)}, z_i^{(2)} | \pi'_{1k}, \Sigma_{il})}{\sum_{k=1}^{K} \pi'_{1k} f(z_i^{(1)}, z_i^{(2)} | \pi'_{1k}, \Sigma_{il})},$$
    $$i = 1, \ldots, n, l = 1, \ldots, K.$$  

  - **M Step:**
    $$\pi'_{1k} = \frac{1}{m + \beta_0}$$
    $$\Sigma_{il} = \frac{\sum_{i=1}^{m} b_{il} z_i^T z_i}{\sum_{i=1}^{m} b_{il}} - 1, l = 1, \ldots, K$$

- until $\pi_1$ and $\Sigma$ converge

- **Jlfdr-thresholding:**
  - Initialize $t(q) = 0$
  - Calculate Jlfdr for each SNP using Eq. (6) with inferred $\pi_1$ and $\Sigma$
  - Sort calculated Jlfdr in ascending order
  - for $a = 1$ to $m$
    - Calculate Fdr(R$_a$) using Eq. (8).
    - if Fdr(R$_a$) $\geq q$, then
      - $t(q)$ = Jlfdr$_{a-1}$; break
    - Output: the SNPs with Jlfdr $\leq t(q)$

There are two tuning parameters in our method—the number of mixture components in the associated SNPs $K$ and the penalization parameter $\beta_0$. Commonly, $K$ has little effect on the fitted distribution of $Z$, which is denoted by $f(Z)$. The increase of $K$ to a well-fitted model will usually result in adding duplicate components or novel components with small proportion. Since Jlfdr(z) = $\pi_0 f_0(z)/f(z)$, $K$ has also little effect on Jlfdr. We use 5-fold cross-validation to select the best set of parameters when $K$ ranges from one to three and $\beta_0$ ranges from 0.1 to 0.5 m. The performance of each fold is evaluated by using the centered $T$-statistic described in Duong et al. (2012). This is a discrepancy measure between the fitted distribution and the empirical distribution of $Z$.

Duong et al. (2012) proposed a kernel-density-based estimation method to evaluate the test statistic. In order to reduce the computational cost of calculating the centered $T$-statistic in multidimensional space, we randomly sample 1000 observations to obtain the empirical distribution of $Z$. The set of parameters with the minimum average $T$-statistic in the cross-validation is selected.

### 2.4 Relationship between the Jlfdr-based method and meta-analysis methods

In the Supplementary Note, we present different forms of rejection regions using the Jlfdr-based method under the Gaussian mixture model, the fixed-effects meta-analysis method and the random-effects meta-analysis method, respectively.

If no heterogeneity exists between studies, the rejection region of the Jlfdr-based method is asymptotically

$$\mathcal{R} = \{z : |z^T x| \geq C\},$$  

(10)

where $x = (\sqrt{n} \bar{x}^{(2)} - \bar{x}^{(1)})^T$ and $C$ is a constant determined by $	ext{Fdr}(\mathcal{R}) = q$. This region coincides with the rejection region of the fixed-effects meta-analysis method. Hence, the Jlfdr-based method and the fixed-effects meta-analysis method will have the same performance when no heterogeneity exists. If heterogeneity exists between studies, their rejection regions are different. In this case, the Jlfdr-based method is more powerful than the fixed-effects meta-analysis method according to Theorem 1.

In any case, the rejection region determined by the Jlfdr-based method and the random-effects meta-analysis method are different. The Jlfdr-based method is more powerful than the random-effects meta-analysis method.

### 3 Results

#### 3.1 Simulation experiments

We use simulation experiments to demonstrate that the Jlfdr-based method is more powerful than the commonly used meta-analysis methods in analyzing summary statistics from multiple GWASs.

In our simulation experiments, we fix the sample size at 10 000 in study 1. We conduct experiments with different sample sizes of 5000, 10 000 and 15 000 in study 2. The sample size ratios $n^{(2)}/n^{(1)}$ are 0.5, 1 and 1.5 correspondingly. The individual numbers in the control group and case group are the same in both studies, and the number of SNPs is $m = 1 \times 10^5$. We simulate the minor allele frequency of each SNP according to uniform distribution $U(0.05, 0.5)$. The proportion of the associated SNPs is 5%. For associated SNPs, the expected log-odds ratio $\mu_0$ in each study is simulated according to the following model:

$$\mu_0/\mu_1 \sim N(\mu, \tau^2)$$  

(11)

where $\tau^2 = 0.04$. In the homogeneous setting, $\tau = 0$. In the heterogeneous setting, $\tau = 0.5$. For non-associated SNPs, the expected log-odds ratio $\mu_0$ is 0. The prevalence of the disease is 1%. We use the log-odds ratio test to detect associations in our experiments.

We use the Jlfdr-based method, the fixed-effects meta-analysis method and the random-effects meta-analysis method to jointly analyze summary statistics from study 1 and study 2. The Fdr is controlled at $q = 5 \times 10^{-5}$. In the fixed-effects meta-analysis and the random-effects meta-analysis, we use the one-dimensional mixture method Muralidharan (2010) to control the Fdr at $q$. 

Downloaded from https://academic.oup.com/bioinformatics/article-abstract/33/4/500/2726128 by guest on 29 April 2019
In the homogeneous setting ($\tau = 0$), each SNP shares the same expected effect size between the two studies. Figure 2 presents the average empirical power and the average Fdp in ten experimental runs using different methods. The average Fdp is well controlled in all methods. This verifies the previous statement that the average empirical power and the average Fdp in the homogeneous setting ($\tau = 0$) are almost the same in the Jlfdr-based method and the fixed-effects meta-analysis method. The subtle differences are due to random initial choices of the EM-algorithm and the Fdr approximations used in Eq. (8). This verifies the previous statement about the equivalence between the Jlfdr-based method and the fixed-effects meta-analysis method in the homogeneous setting. In Supplementary Table S1, we present the average empirical power and the average Fdp using the Jlfdr-based method with different $K$ and $\beta_0$. The Jlfdr-based method is not sensitive to the parameters.

In the heterogeneous setting ($\tau = 0.5$), the expected effect sizes of each SNP vary between studies. Figure 3 plots the discovered associations using the Jlfdr-based method and the fixed-effects meta-analysis method in one run when $n^{(2)} = 10\,000$. Although the Jlfdr-based method missed some associations detected by the fixed-effects meta-analysis method, it identifies more associations than the meta-analysis method. We ran the simulation experiments ten times for the sample size ratio $n^{(2)}/n^{(1)} = 0.5, 1$ and $1.5$. Figure 4 shows the average empirical power and the average Fdp. The average Fdp using all three methods are about $q = 5 \times 10^{-5}$ in all sample size ratio settings. From the figure, we can see that the Jlfdr-based method can achieve higher power than the other methods when controlling Fdr at the same threshold. In Supplementary Table S2, we show the average empirical power and the average Fdp using the Jlfdr-based method with different $K$ and $\beta_0$. The Jlfdr-based method is not sensitive to the parameters.

In order to check the robustness of our method under model misspecification, we also apply the Jlfdr-based method and meta-analysis methods to simulate data with the following effect size distribution:

$$\mu|H_1 \sim t_{5,0.2},$$

where $t_{5,0.2}$ is a scaled t-distribution with degree of freedom 5 and scaling factor 0.2. In Supplementary Figures S1 and S2, we show the average empirical power and the average Fdp in the homogeneous and heterogeneous settings, respectively. When controlling Fdr at the same level, the Jlfdr-based method and the fixed-effects meta-analysis method have almost the same average empirical power in the homogeneous setting. The Jlfdr-based method can achieve higher power in the heterogeneous setting.

### 3.2 Real data applications

#### 3.2.1 SCZ data from PGC

We jointly analyze the summary statistics from schizophrenia (SCZ) studies conducted by the Psychiatric Genomics Consortium (PGC). The summary statistics from two SCZ studies, Sweden + SCZ1 (Ripke et al., 2013) and SCZ2 (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), are available from the PGC. Sweden + SCZ1 is a large-scale meta-analysis of Swedish and mixed-European ancestry individuals that comprises 13 833 schizophrenia cases and 18 310 controls in the analysis. We use it as Study
Multiple genome-wide association studies

3.2.2 SLE data from dbGaP

We conduct summary-statistics-based joint analysis in systemic lupus erythematosus (SLE) data from phs000122.v1.p1 and phs000216.v1.p1 in dbGaP (Mailman et al., 2007; Tryka et al., 2014). We use the study phs000122.v1.p1, in which there are 1311 SLE cases and 3340 controls, as Study 1, and we use the study phs000216.v1.p1, in which there are 706 cases and 353 controls, as Study 2. The individuals in the first study are all North Americans of European descent, and those in the second study are all females of European ancestry. We use the following quality control procedures for both studies:

1. Missing data control: The SNPs with a missing data rate larger than 1% are discarded.
2. Minor allele frequency control: The SNPs with minor allele frequency less than 0.05 in either case group or control group are discarded.
3. Hardy-Weinberg equilibrium control: In the Hardy-Weinberg equilibrium test, the SNPs with P-values less than 0.001 in either case group or control group are discarded.
4. Homogeneity control: In the homogeneity test, SNPs with P-values less than 0.01 are discarded.

After the quality control steps, there are $m = 195,318$ autosomal SNPs remaining.

We use $q = 5 \times 10^{-5}$ as the Fdr threshold in all analyses. In Supplementary Table S5, we present the number of identified SNPs using the Jlfdr-based method with different $K$ and $\beta_0$. We use 5-fold cross-validation to select parameters, and the selected values are $K = 2$ and $\beta_0 = 0.2m$. We adopt the kernel-density-based global two-sample comparison test (Duong et al., 2012) to measures the goodness of fit of the Gaussian mixture model, the P-value is 0.20 (We randomly sample 1000 observations in calculating the empirical distribution of $Z$ due to the heavy computational burden).

Figure 5(a) plots the discovered associations using the Jlfdr-based method and the fixed-effects meta-analysis method. The Jlfdr-based method identifies more associations. Table 3 shows the numbers of discovered associations and the rejection criteria of the different analysis methods. Besides the loci discovered by meta-analysis methods, there are 17 novel loci discovered by the Jlfdr-based method. Each locus is separated by at least 500 kilobases (kb) or a weak linkage disequilibrium ($r^2 < 0.1$). The SNPs showing the most significant association with SCZ in these novel loci are presented in Supplementary Table S4.

1. SCZ2 is a larger-scale meta-analysis that comprises 36 989 schizophrenia cases and 113 075 controls. The analysis includes the individuals which have been analyzed in Sweden + SCZ1. By using the following inverse meta-analysis formula, we obtain the summary statistics from the meta-analysis comprising the individuals only be analyzed in SCZ2. The formula is

$$ z^{(2)} = \frac{z_w / \sigma_w - z^{(1)} / \sigma^{(1)}}{1 / (\sigma_w)^2 - 1 / (\sigma^{(1)})^2}. \tag{13} $$

We use $z^{(2)}$ as the summary statistics of Study 2. We remove the SNPs with $P$-value $< 0.01$ in the test of homogeneity. After that, there are $m = 8,157,410$ SNPs remaining.

We use the Jlfdr-based method, the fixed-effects meta-analysis method and the random-effects meta-analysis method to jointly analyze the summary statistics from two studies. The Fdr is controlled at $q = 5 \times 10^{-5}$. We adopt the one-dimensional mixture method to control the Fdr at $q$ in meta-analysis methods. In Supplementary Table S3, we show the number of identified SNPs using the Jlfdr-based method with different $K$ and $\beta_0$. Here we use 5-fold cross-validation to select parameters, and the selected values are $K = 3$ and $\beta_0 = 0.2m$. We adopt the kernel-density-based global two-sample comparison test (Duong et al., 2012) to measures the goodness of fit of the Gaussian mixture model, the P-value is 0.20 (We randomly sample 1000 observations in calculating the empirical distribution of $Z$ due to the heavy computational burden).

Figure 5(a) plots the discovered associations using the Jlfdr-based method and the fixed-effects meta-analysis method. The Jlfdr-based method identifies more associations. Table 3 shows the numbers of discovered associations and the rejection criteria of the different analysis methods. Besides the loci discovered by meta-analysis methods, there are 17 novel loci discovered by the Jlfdr-based method. Each locus is separated by at least 500 kilobases (kb) or a weak linkage disequilibrium ($r^2 < 0.1$). The SNPs showing the most significant association with SCZ in these novel loci are presented in Supplementary Table S4.
Table 4. The rejection criterion and the number of identified associations in SLE data from dbGaP

<table>
<thead>
<tr>
<th>Method</th>
<th>Rejection Criterion</th>
<th>#{SNPs}</th>
<th>#{loci}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jlfdr-based method</td>
<td>Jlfdr(z) ≤ 5.853 × 10^{-4}</td>
<td>107</td>
<td>14</td>
</tr>
<tr>
<td>Fixed-effects meta-analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jlfdr(z) ≥ 5.570</td>
<td>94</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Random-effects meta-analysis</td>
<td>Jlfdr(z) ≥ 5.617</td>
<td>53</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 5. The rejection criterion and the number of identified associations in BMI data from the GIANT

<table>
<thead>
<tr>
<th>Method</th>
<th>Rejection Criterion</th>
<th>#{SNPs}</th>
<th>#{loci}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jlfdr-based method</td>
<td>Jlfdr(z) ≤ 3.535 × 10^{-4}</td>
<td>2839</td>
<td>209</td>
</tr>
<tr>
<td>Fixed-effects meta-analysis</td>
<td>Jlfdr(z) ≥ 5.367</td>
<td>2618</td>
<td>192</td>
</tr>
<tr>
<td>Random-effects meta-analysis</td>
<td>Jlfdr(z) ≥ 5.430</td>
<td>2108</td>
<td>152</td>
</tr>
</tbody>
</table>

\(z_{MF}\) and \(z_{MR}\) are the combined z-values in the fixed-effects meta-analysis and random-effects meta-analysis, respectively. #{SNPs} and #{loci} are the numbers of identified SNPs and loci, respectively.

3.2.3 BMI data from GIANT

We jointly analyze summary statistics from body mass index (BMI) studies conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium (Locke et al., 2015). We use the joint GWAS and metabochip meta-analysis of 152 893 European men as Study 1, and we use the joint GWAS and metabochip meta-analysis of 171 977 European women as Study 2. There are \(m = 2\) 466 338 autosomal SNPs passing the homogeneity control (P-value > 0.01).

We use \(q = 5 \times 10^{-5}\) as the Fdr threshold in all analyses. In Supplementary Table S7, we present the number of identified SNPs using the Jlfdr-based method with different \(K\) and \(b_0\). The selected values of parameters by using cross-validation are \(K = 1\) and \(b_0 = 0.1\) m. In the goodness-of-fit test of the Gaussian mixture model, the P-value is 0.26. Figure 5(c) plots the associations discovered by the Jlfdr-based method and the fixed-effects meta-analysis method. The Jlfdr-based method discovers more associations than meta-analysis methods. Table 5 shows the number of discovered associations and the corresponding rejection criterion of each method. There are 22 novel loci discovered by the Jlfdr-based method. The SNPs showing the most significant associations in these novel loci are listed in Supplementary Table S8.

3.2.4 WHRadjBMI data from GIANT

We conduct joint analysis in waist-to-hip ratio after adjusting for BMI (WHRadjBMI) studies from GIANT consortium (Shungin et al., 2015). We use the joint GWAS and metabochip meta-analysis of 93 480 European men as Study 1, and we use the joint GWAS and metabochip meta-analysis of 116 742 European women as Study 2. There are \(m = 2\) 127 324 autosomal SNPs passing the homogeneity control (P-value > 0.01).

In Supplementary Table S9, we present the number of identified SNPs using the Jlfdr-based method with different \(K\) and \(b_0\). The selected values of parameters by using cross-validation are \(K = 2\) and \(b_0 = 0.1\) m. In the goodness-of-fit test of the Gaussian mixture model, the P-value is 0.17. Figure 5(d) highlights the associations discovered by the Jlfdr-based method and the fixed-effects meta-analysis method. The Jlfdr-based method identifies more associations than meta-analysis methods when controlling Fdr at the same level \(q = 5 \times 10^{-5}\). Table 6 shows the number of discovered associations and the corresponding rejection criterion of each method. Besides the loci discovered by meta-analysis methods, there are four novel loci discovered by the Jlfdr-based method. The details of the most significant SNPs in these loci are listed in Supplementary Table S10.

Table 6. The rejection criterion and the number of identified associations in WHRadjBMI data from the GIANT

<table>
<thead>
<tr>
<th>Method</th>
<th>Rejection Criterion</th>
<th>#{SNPs}</th>
<th>#{loci}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jlfdr-based method</td>
<td>Jlfdr(z) ≤ 4.835 × 10^{-4}</td>
<td>459</td>
<td>47</td>
</tr>
<tr>
<td>Fixed-effects meta-analysis</td>
<td>Jlfdr(z) ≥ 5.664</td>
<td>413</td>
<td>43</td>
</tr>
<tr>
<td>Random-effects meta-analysis</td>
<td>Jlfdr(z) ≥ 5.805</td>
<td>176</td>
<td>26</td>
</tr>
</tbody>
</table>

4 Discussion

All the summary-statistics-based joint analysis methods described above use the theoretical null distribution N(0, 1) as the distribution of z-values under the null hypothesis. There is an argument that the theoretical null distribution is not appropriate to describe most z-values due to correlation across SNPs, correlation across individuals, confounding effects (Efron, 2004). To address this issue, Efron advocates to use an empirical null distribution instead of the theoretical null distribution. Our Jlfdr-based method can easily accommodate the empirical null distribution by assuming N(0, I + \(\Sigma_0\)) as the null distribution of \(Z\). Here \(\Sigma_0\) is a matrix which needs to be estimated in the EM-algorithm. In Supplementary Figure S3 and Supplementary Table S11-S14, we present the experimental results of the above four datasets by using empirical nulls. The numbers of identified associations are smaller than the numbers using theoretical null. However, the Jlfdr-based method can still identify more associations than meta-analysis methods when we use empirical nulls to control Fdr in all methods.

Both the Jlfdr-based method and the meta-analysis methods jointly analyze summary statistics from multiple GWASs. Meta-analysis methods collapse the test statistics of all studies into a weighted average value for each SNP, which is simpler than the Jlfdr-based method. When no heterogeneity exists between studies, the Jlfdr-based method will degenerate to the fixed-effects meta-analysis method. This can be understood by the fact that there is no information loss during the collapsing when all studies are homogeneous. When heterogeneity exists between studies, however, the Jlfdr-based method can achieve higher power than the fixed-effects meta-analysis method. This is understandable as information about heterogeneity is lost during collapse when using the meta-analysis method. Since heterogeneity widely exists in most cases, we suggest to use the Jlfdr-based method instead of meta-analysis methods to jointly analyze summary statistics from multiple GWASs.

This paper proves that the Jlfdr-based method is the most powerful summary-statistics-based joint analysis method when the underlying distribution of the test statistics is known. In reality, we only know the theoretical distribution under a null hypothesis. The distribution under alternative hypotheses is usually unknown. Hence, in the implementation of the Jlfdr-based method, we assume test statistics follow a unimodal Gaussian mixture model. Then we use the EM-algorithm to infer parameters in the mixture model. The Gaussian mixture model is a semi-parametric model and provides great flexibility and precision in modelling underlying data. We can use it to approximate arbitrary unimodal distribution to any fidelity with appropriate number of mixtures (Park and Sandberg, 1991). However, violation of the model assumptions and inaccuracy of parameters estimation will decrease the performance of the Jlfdr-based
method. We may further improve the Jlfdr-based method by relaxing the model assumption.

We assume an independence between SNPs in the Gaussian mixture model. In other words, we only consider marginal information of each SNP in our method. However, correlations between nearby SNPs often exist, which is known as linkage disequilibrium. Recently, Heggeseth and Jewell (2013) shows that the EM algorithm used in the Gaussian mixture model has a certain robustness to misspecified dependency relationship. Even so, we may further improve the Jlfdr-based method by taking advantage of the dependency information between SNPs.

Funding
This paper was partially supported by a grant under the Theme-based Research Scheme (project T12-402/13N) of the Hong Kong Research Grant Council (RGC).

Conflict of Interest: none declared.

References