COMPUTATIONAL
AND STRUCTURAL
BIOTECHNOLOGY
J O U R N A L

Evaluating statistical significance in a *meta*-analysis by using numerical integration

Yin-Chun Lin [a], Yu-Jen Liang [a], Hsin-Chou Yang [a,b,c,d,*]

[a] *Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan*
[b] *Biomedical Translation Research Center, Academia Sinica, Taipei 11529, Taiwan*
[c] *Institute of Statistics, National Cheng Kung University, Tainan 70101, Taiwan*
[d] *Institute of Public Health, National Yang-Ming University, Taipei 11221, Taiwan*

ARTICLE INFO

ABSTRACT

Meta-analysis is a method for enhancing statistical power through the integration of information from multiple studies. Various methods for integrating p-values (i.e., statistical significance), including Fisher's method under an independence assumption, the permutation method, and the decorrelation method, have been broadly used in bioinformatics and computational biotechnology studies. However, these methods have limitations related to statistical assumption, computing efficiency, and accuracy of statistical significance estimation. In this study, we proposed a numerical integration method and examined its theoretical properties. Simulation studies were conducted to evaluate its Type I error, statistical power, computational efficiency, and estimation accuracy, and the results were compared with those of other methods. The results demonstrate that our proposed method performs well in terms of Type I error, statistical power, computing efficiency (regardless of sample size), and statistical significance estimation accuracy. P-value data from multiple large-scale genome-wide association studies (GWASs) and transcriptome-wise association studies (TWASs) were analyzed. The results demonstrate that our proposed method can be used to identify critical genomic regions associated with rheumatoid arthritis and asthma, increase statistical significance in individual GWASs and TWASs, and control for false-positives more effectively than can Fisher's method under an independence assumption. We created the software package Pbine, available at GitHub (https://github.com/Yinchun-Lin/Pbine).

1. Introduction

Fisher's p-value combination method [1], which integrates statistical significance (i.e., p-value p_i) from many statistical hypothesis tests was originally proposed to examine a joint null hypothesis (i.e., an intersection of the individual null hypotheses) in a *meta*-analysis [2]. Test statistic of Fisher's method is defined as negative two times the summation of log-transformed p-values (i.e., $F = -2\sum_i \ln(p_i)$) [1]. This method has been broadly applied in bioinformatics and computational biotechnology studies, such as

the *meta*-analysis of genome-wide association study (GWAS) [3–5] and that of transcriptome-wide association study (TWAS) [6–8].

If all the p-values $\{p_i, i = 1, \cdots, K\}$ follow Uniform (0,1) distribution under a null hypothesis independently, test statistic F follows a chi-squared distribution with $2K$ degrees of freedom. As such, the exact p-value (p_F) of F can be derived. However, when the p-values are correlated, the mathematical derivation for the sampling distribution of F becomes intractable. Remarkably, ignorance of the correlation of p-values will inflate false-positives in the subsequent statistical inference [9].

The permutation procedure [10], which involves non-parametric resampling without replacement based on a set of observed data, has been applied to generate a null distribution and calculate an empirical p-value (p_P) of F when the independence assumption of p-values is violated. The permutation procedure is simple in concept and robust to various correlation structures of p-values. However, this method has some limitations.

For instance, in a GWAS with a large sample size, which is used to evaluate the phenotype–genotype relationship, permutations over the phenotype status of study samples require intensive computation. Permutations over single nucleotide polymorphisms (SNPs) may distort the inherent structure of SNPs (i.e., linkage disequilibrium). In addition, numerous permutations and intensive computation are required for a correction for multiple testing in GWASs and TWASs [11,12].

The permutation procedure requires raw data (only p-value data itself cannot perform permutations), which are not always available. For instance, in a *meta*-analysis of GWAS, the p-values of single-locus association tests from public genomic databases (e.g., GWAS Catalog [13] and GWAS Central [14]) are available and can be combined to infer the genetic association between SNPs and a phenotype of interest. However, no raw genotype and phenotype data are provided in GWAS Catalog and GWAS Central.

Without the need for raw data, a decorrelation procedure [15] transforms dependent p-values into independent p-values only on the basis of p-value data and the correlation structure of the p-values. The distribution F can thus be derived and computed rapidly. However, in this study, we demonstrated that the order of the combination of p-values influences the results (the order-noninterchangeable property). For example, if $p_1 = 0.9970$ and $p_2 = 0.1903$ are combined and their correlation coefficient $r = 0.5$, then the decorrelation procedure provides a p-value of F (p_D) = 0.0297; however, if the order of the combination of p-values is reversed, then $p_D = 0.5060$.

Considering the permutation and decorrelation methods' limitations, we suggested a numerical integration method (p_N) to evaluate the statistical significance of F. Herein, the theoretical properties of a p-value combination are examined. Type I error, statistical power, computational efficiency, and estimation accuracy were evaluated through simulation studies and compared with those of Fisher's method under an independence assumption (p_F), permutation (p_P), and decorrelation (p_D). Real-world examples of *meta*-analyses of GWASs and *meta*-analyses of TWASs are given. R codes Pbine are provided in GitHub at https://github.com/Yinchun-Lin/Pbine.

2. Methods

2.1. Proposed method

Let $\mathbf{R} = \{p_i, i = 1, \cdots, K\}$ denote a list of dependent p-values from K hypothesis tests with a correlation $corr(\mathbf{R}) = \Sigma$. Because the correlation is approximately invariant under monotone transformations [15], $\mathbf{R}$ satisfies the following transformation $\mathbf{T}$: $\mathbf{R} = 1 - \Phi\{\mathbf{C}\Phi^{-1}(1 - \mathbf{R}^{\#})\}$, where $\mathbf{C}$ is the Cholesky factor of Σ (i.e. $\Sigma = \mathbf{C}\mathbf{C}^{\mathrm{T}}$), $\mathbf{R}^{\#}$ is a list of independent p-values, and Φ denotes the cumulative distribution function of a standard normal random variable. Let $\{p_i^*, i = 1, \cdots, K\}$ be the p-values calculated from the raw data observations for testing K null hypotheses. In our method, the p-value of F is calculated as follows:

$$p_N = \int_D J_{T^{-1}},$$

where $J_{T^{-1}}$ is the joint probability density function of $\mathbf{R}$ (**Text S1**), D is the region enclosed by $0 < p_i < 1, \forall i$ and $\prod p_i < \prod p_i^*$. Noted that p_N is fixed for any permutation on $\{p_i^*, i = 1, \cdots, K\}$.

For reduced case $K = 2$ with correlation $0 < r < 1$, then $\mathbf{C} = \begin{pmatrix} 1 & 0 \\ r & \sqrt{1-r^2} \end{pmatrix}$. The p-value of F is derived (**Text S1**) as follows:

$$p_N = \iint_D \frac{1}{\sqrt{1-r^2}} \Phi' \left(\frac{\Phi^{-1}(1-v) - r\Phi^{-1}(1-u)}{\sqrt{1-r^2}} \right)$$
$$\times \left(\Phi^{-1} \right)'(1-v)d(u,v), \tag{1}$$

where $D = \{(u,v)|0 < u, v < 1, uv < p_1^* p_2^*\}$ and Φ' is the derivative of a cumulative distribution function of a standard normal variate.

A logarithm transformation of p_i's was considered to prevent integration on curvilinear region D. p_N in **Eq. (1)** can be rewritten as follows:

$$p_N = 1 - \iint_{D'} \frac{1}{\sqrt{1-r^2}} \Phi' \left(\frac{\Phi^{-1}(1-e^{v'}) - r\Phi^{-1}(1-e^{u'})}{\sqrt{1-r^2}} \right)$$
$$\times \left(\Phi^{-1} \right)'(1-e^{v'})e^{u'}e^{v'}d(u',v')$$

where $(u', v') = (\ln(u), \ln(v))$ and $D' = \{(u', v')|u', v' < 0, u' + v' > \ln(p_1^*) + \ln(p_2^*)\}$.

If correlation of two p-values is $r = 0$, p_N in **Eq. (1)** can be reduced to Fisher's method as follows:

$$p_N = p_F = p_1^* p_2^* - p_1^* p_2^* \ln(p_1^* p_2^*) \tag{2}$$

For $K = 2$, the detailed derivation of $J_{T^{-1}}$ (**Text S1**), the symmetric behavior and limiting behavior of $J_{T^{-1}}$ (**Text S2**), and a comparison of the p-value of our proposed method (p_N) with the individual p-values (**Text S3** and **Fig. S1**) are provided.

2.2. Simulation studies

To evaluate the Type I error, power, and computation time of the p-value combination methods, bivariate linear regression models were applied to generate data in the simulation studies. Model parameters were assigned or estimated following the purposes of the simulation study. The p-values of our numerical integration method (p_N), the Fisher's method under an independence assumption (p_F), and the decorrelation method (p_D) were calculated and compared with the benchmark method – the permutation method (p_P).

2.2.1. Evaluation of Type I error

A bivariate linear regression model was applied to generate two phenotypes (y_{ng}, z_{ng}) and gene expression (x_{ng}) data in the simulation study as follows:

$$\begin{pmatrix} y_{ng} \\ z_{ng} \end{pmatrix} = \begin{pmatrix} \beta_j x_{ng} \\ \gamma_j x_{ng} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ng} \\ \varphi_{ng} \end{pmatrix}, n = 1, \cdots, N, g = 1, \cdots, G \tag{3}$$

where N denotes the total number of samples and G indicates the total number of genes. Random error terms ε_{ng} and $\varphi_{ng} = w\varepsilon_{ng} + (1-w)\epsilon_{ng}$ have correlation coefficient $corr(\varepsilon_{ng}, \varphi_{ng}) = w$. We assumed that ε_{ng} and ϵ_{ng} would be independent white noises following a standard normal distribution. Gene expression x_{ng} was independently generated from Uniform(0, 1).

Here, we considered the regression coefficient under the null hypothesis: $\beta_j = \gamma_j = 0$; sample size $N = 500$; gene number $G = 2,000$; simulation replications $K = 2,000$. Given a specified w, we generated the phenotypes and gene expression data. In each of the simulated datasets, p-values $\mathbf{p}_y$ and $\mathbf{p}_z$ were obtained by testing null hypotheses $H_{0g}^{\beta}: \beta_g = 0$ and $H_{0g}^{\gamma}: \gamma_g = 0$, respectively, through Student's t test.

The results demonstrate that the methods control Type I error well for uncorrelated p-values, i.e., $corr(\mathbf{p}_y, \mathbf{p}_z) = 0$ (Fig. 1A). For the positively correlated p-values ($\mathbf{p}_y, \mathbf{p}_z$), Fisher's method (red

line) under an independence assumption (p_F) exhibited an inflated Type I error particularly for an increased correlation of p-values. Our method p_N and the decorrelation procedure p_D exhibited a Type I error similar to that of the benchmark method p_P; the false-positive was 0.05–0.06. This indicates that p-value dependency must be considered when the sampling distribution of a p-value combination is derived. Therefore, our proposed method controls Type I errors well.

2.2.2. Evaluation of statistical power

A bivariate linear regression model was applied to generate two phenotypes (y_{ng}, z_{ng}) and gene expression (x_{ng}) data in the simulation study as follows:

$$\begin{pmatrix} y_{ng} \\ z_{ng} \end{pmatrix} = \begin{pmatrix} \beta_j x_{ng} \\ w\beta_j x_{ng} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ng} \\ \varphi_{ng} \end{pmatrix}, n = 1, \cdots, N, g = 1, \cdots, G \quad (4)$$

where N indicates the total number of samples and G indicates the total number of genes. That is, $z_{ng} = wy_{ng} + (1-w)\epsilon_{ng}$ and $\varphi_{ng} = w\varepsilon_{ng} + (1-w)\epsilon_{ng}$, and $corr(y_{ng}, z_{ng}) = corr(\varepsilon_{ng}, \varphi_{ng}) = w$. We assumed that random error terms ε_{ng} and ϵ_{ng} would follow an independent standard normal random distribution individually. Gene expression x_{ng} was generated from Uniform$(0,1)$ independently. Here, we considered $\beta_j = 0.6$; $w = 0.1, 0.3, 0.5, 0.7, 0.9$; sample size $N = 500$; number of genes $G = 2,000$; simulation replications $K = 2,000$. P-values $\boldsymbol{p}_y$ and $\boldsymbol{p}_z$ were obtained by testing null hypotheses $H_{0g}^\beta : \beta_g = 0$ and $H_{0g}^\gamma : \gamma_g = 0$, where $\gamma_g = w\beta_g$. The relationship between $corr(\varepsilon_{ng}, \varphi_{ng}) = w$ and $corr(\boldsymbol{p}_y, \boldsymbol{p}_z) = r$ is discussed (**Text S4** and **Fig. S2**).

The results demonstrated that the four methods have similar power and that the power increases with r (Fig. 1**B**). p_F has the highest power, particularly for a higher r; however, the high power is accompanied by an inflated Type I error as mentioned in the pre-

vious section (Fig. 1**A**). p_D has the lowest power and deviates from the benchmark p_P, particularly for a high r. The proposed method p_N has high power similar to that of the benchmark p_P.

2.2.3. Evaluation of computation time

The simulation model for an evaluation of Type I error was then applied. We compared the computation time between the proposed method (p_N) and the benchmark method (p_P) with various sample sizes ($N = 3,000, 5,000,$ and $7,000$), gene numbers ($G = 10,000, 20,000,$ and $30,000$), correlation coefficients ($w = 0.1, 0.5,$ and 0.9), and permutation times ($K = 2,000$–$20,000$, with increments of $2,000$). The results show that the computation time of the proposed method (p_N) is unchanged with the sample size, number of genes, and correlation coefficient (Fig. 1**C**). The computation time for the permutation method (p_P) increases with sample size, number of genes, and permutation time, but it did not change as the correlation coefficient increased; it required approximately 2,500 h to compute 30,000 pairs of p-value combinations. Our proposed method was more computationally efficient, requiring < 10 h for computation. Because Fisher's method under an independence assumption (p_F) and the decorrelation method (p_D) do not involve complex re-sampling or integration procedures, they are computationally efficient.

2.2.4. Evaluation of estimation accuracy

We applied the model for an evaluation of Type I error with $G = 1,000$, $N = 500$, and $K = 5,000$, and compared the estimation accuracy of the three p-value combination methods p_N, p_F, and p_D with that of the benchmark permutation p_P. Correlation coefficients $corr(\boldsymbol{p}_y, \boldsymbol{p}_z) = r = 0.1, 0.5, 0.9$ were considered. The results demonstrate that p_N, p_F, and p_D deviate from the benchmark p_P as $corr(\boldsymbol{p}_y, \boldsymbol{p}_z) = r$ increases (Fig. 1**D**–1**F**). In addition, p_D exhibits a certain proportion of outliers (Fig. 1**D**–1**F**). The biased estimation can be explained by the order-noninterchangeable property (**Text**
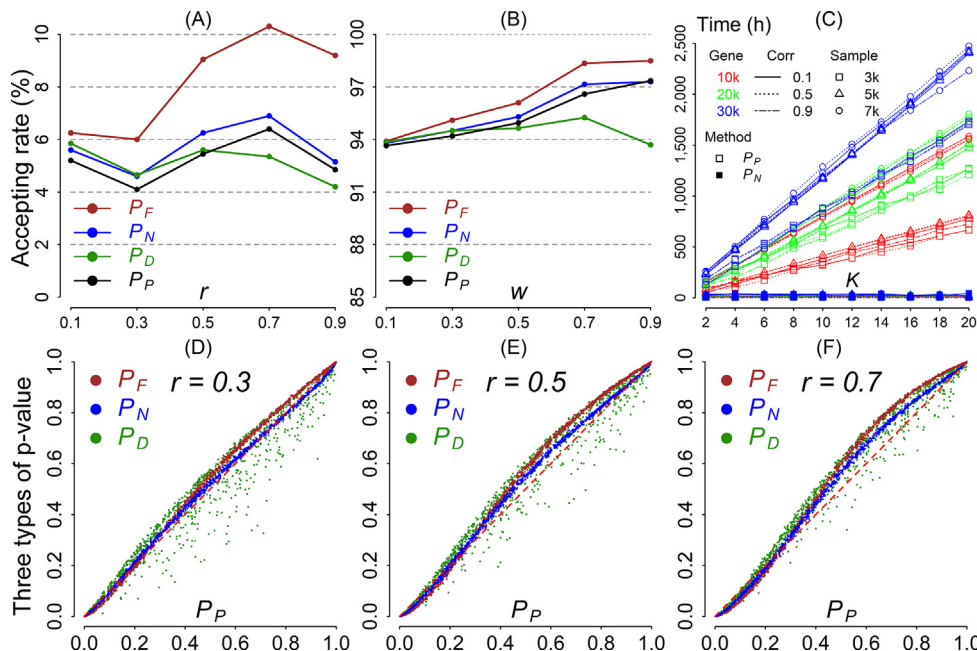
Fig. 1. **(A) Type I error for the p-value combination methods.** The x-axis represents correlation $corr(\boldsymbol{p}_y, \boldsymbol{p}_z) = r$. The y-axis represents Type I error. **(B) Statistical power for the p-value combination methods.** The x-axis represents correlation $corr(\varepsilon_{ng}, \varphi_{ng}) = w$. The y-axis represents statistical power. **(C) Computation time of the proposed method (p_N) and the permutation method (p_P).** The x-axis indicates number of thousand permutations (K). The y-axis indicates computation time. Red, green, and blue lines denote gene sizes G of 10,000, 20,000, and 30,000, respectively. The squares, triangles, and circles denote sample sizes N of 3,000, 5,000, and 7,000, respectively. Solid, dotted, and dot-dashed lines represent correlation coefficients r of 0.1, 0.5, and 0.9, respectively. **(D)–(F) Estimation accuracy of p-value.** The x-axis represents the p-values of the benchmark method (p_P). The y-axis indicates the p-values of the other p-value combination methods: p_F, brown; p_N, blue; p_D, green). The results based on correlation coefficients $r = 0.3, 0.5,$ and 0.7 are arranged from left to right.

S5) and non-uniformity property (**Fig. S3**) of p_D. Compared with p_F and p_D, the proposed method (p_N) is closest to the benchmark p_P. Thus, the proposed method provides a more accurate estimate than do p_F and p_D.

3. Real data applications

3.1. Meta-GWAS for rheumatoid arthritis

This *meta*-analysis identified SNPs associated with rheumatoid arthritis on the basis of two large-scale population-based GWASs – The North American Rheumatoid Arthritis Consortium (NARAC) data ($N_{case} = 868, N_{control} = 1,194$)[16] and Wellcome Trust Case Control Consortium (WTCCC) data ($N_{case} = 1,999, N_{control} = 3,002$) [17]. In each of the two GWASs, a logistic regression analysis with covariate adjustment for sex and SNP coding based on an additive genetic model was performed to examine the genetic associations between rheumatoid arthritis disease status and individual SNP markers. At each SNP, p-values based on NARAC and WTCCC data were obtained separately.

We applied Fisher's (p_F) and our (p_N) methods to combine the two p-values at each SNP locus in the NARAC and WTCCC data. There were 4,963 statistical tests (because of 4,963 SNPs on chromosome 6) in this *meta*-GWAS. Bonferroni's correction [18] for multiple testing was performed to obtain adjusted p-values for p_F and p_N separately. The result demonstrate that both the methods could be used to identify the major histocompatibility complex (MHC) region on chromosome 6p21.3 (Fig. 2), which is strongly associated with rheumatoid arthritis [19,20]. Our method ennabled us to identify the SNP rs9391858 (p = 3.495×10^{-6}) truly associated with rheumatoid arthritis [21] (Fig. 2); however, the individual GWASs could not detect this SNP p = 1.020×10^{-5} in NARAC and p = 1.291×10^{-4} in WTCCC; Fig. S4). Fisher's method but not ours identified six false-positive SNPs: rs2394102, rs11752073, rs12697946, rs9394169, rs3818528, and rs3130014 (Fig. 2).

The real data analysis demonstrated that our method can be used to identify crucial genomic regions strongly associated with rheumatoid arthritis, detect some of rheumatoid arthritis-associated loci not detected by individual GWASs, and control for false-positive more efficiently than can Fisher's method.
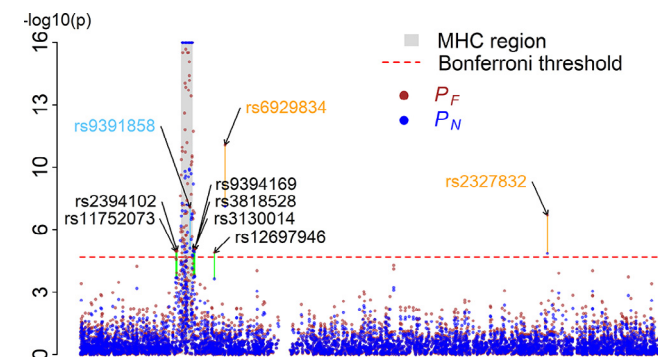
Fig. 2. Manhattan plots for chromosome 6 in the *meta*-GWAS. This *meta*-GWAS contained 4,963 SNPs on chromosome 6. The Fisher's method (p_F) and our method (p_N) were employed. Each point indicates a SNP. The x-axis indicates physical position of a SNP. The y-axis indicates p-value in a scale of −log_{10}. The green lines indicate false-positive events identified by the Fisher's method (p_F) but not by our method (p_N); they involved six SNPs: rs11752073, rs2394102, rs3130014, rs3818528, rs9394169, and rs12697946 (green line). The orange lines indicate false-positive events identified by both of the Fisher's method (p_F) and our method (p_N). The light blue line indicates the SNP known to be associated with rheumatoid arthritis and identified by the Fisher's method (p_F) and our method (p_N), but not by either of the two studies (Fig. S4). The red dashed line indicates the significance level after Bonferroni correction for multiple testing.

3.2. Meta-TWAS for rheumatoid arthritis

This *meta*-analysis was performed to identify genes differentially expressed in patients with rheumatoid arthritis and normal controls of European descent based on the two large-scale TWASs: Eyre et al. ($N_{case} = 13,838, N_{control} = 33,742$) [22] and Stahl et al. ($N_{case} = 5,539, N_{control} = 20,169$) [23]. In each TWAS, the gene-level p-values used to examine the association of rheumatoid arthritis disease status with gene expression were downloaded from webTWAS [24]. However, raw gene expression data were unavailable on the website. When the p-value data were downloaded, Elastic-net was used as a model for transcriptome data prediction in the MetaXcan framework [25].

We applied Fisher's (p_F) and our (p_N) methods to combine the two p-values from the two TWASs. There were 3,175 statistical tests (because of 3,175 genes overlapping between the two studied TWAS datasets) in this *meta*-TWAS. Bonferroni's correction [18] for multiple testing was performed to obtain adjusted p-values for p_F and p_N separately. The results indicate that both methods identified the MHC region as a key genomic region for rheumatoid arthritis (Fig. 3). Fisher's method identified 11 genes outside the MHC region. Except *ANKRD55* was a true-positive, all other 10 genes were false-positively identified. Our method did not have false-positives but failed to detect *ANKRD55* ($p_N = 4.410 \times 10^{-5}$). Our and Fisher's methods did not identify additional genes associated with rheumatoid arthritis detected by individual TWASs; nevertheless, our method demonstrated a more significant signal for several rheumatoid arthritis genes. For instance, in the TWAS of Eyre et al [22], the TWAS of Stahl et al [23], and this *meta*-TWAS, the p-values were p = 1.93×10^{-6}, p = 6.33×10^{-7}, and p = 1.39×10^{-7}, respectively, for *AFF3* on chromosome 2, and p = 3.07×10^{-5}, p = 1.30×10^{-5}, and p = 4.09×10^{-6}, respectively, for *IRF5* on chromosome 7 (**Fig. S5**). Thus, this real data analysis demonstrated the inference of our *meta*-TWAS.

3.3. Meta-TWAS for asthma

We evaluated the performance of the proposed method in combining p-values from more than two studies. We downloaded the gene-level p-values data in four large-scale studies for asthma from webTWAS [24] – Canela-Xandri et al. [26] with $N_{case} = 52,269$ and $N_{control} = 399,995$; Zhu et al. [27] with $N_{case} = 14,085$ and $N_{control} = 76,768$; Zhu et al. [28] with
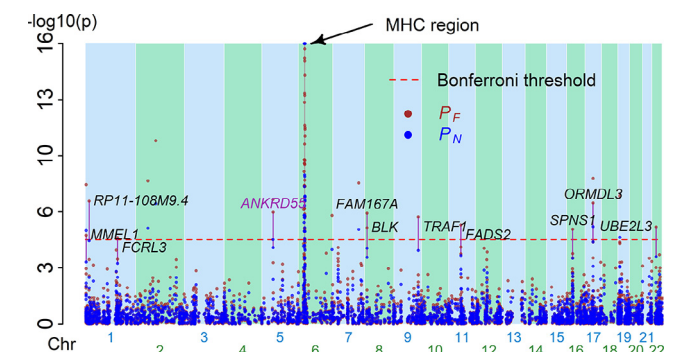
Fig. 3. Manhattan plots for 22 autosomes in the *meta*-TWAS. This *meta*-TWAS contained 3,175 genes overlapping between the two studied TWAS datasets. The Fisher's method (p_F) and our method (p_N) were employed. Each point indicates a gene. The x-axis indicates the physical position of a gene in an autosome. The y-axis indicates p-value in a scale of −log_{10}. The 11 purple lines indicate the genes for which Fisher's method (p_F) reported genetic association but our method (p_N) did not. All genes, *ANKRD55*, were false-positively detected. The red dashed line indicates the significance level after Bonferroni correction for multiple testing.

Table 1

Our method performs better than Fisher's method when combining more than two p-values. Ten asthma related genes and five asthma unrelated gene were analyzed in this *meta*-TWAS. Adjusted p-values of Fisher's method (p_F), our method with equal weights (p_N), and our method with unequal weights (p_N^w) are provided. All the p-values were adjusted using Bonferroni's correction for multiple testing. The numbers marked in bold indicate they are statistically significant.

	Gene name	Chromosome	p_F	p_N^w	p_N
Asthma Related Genes	*HLA-G*	Chr. 6	**0.0000127**	**0.0244885**	**0.0012742**
	ATP6V1G2	Chr. 6	**0.0005959**	**0.0030811**	**0.0168662**
	HLA-DRB5	Chr. 6	**0.0059697**	**0.0386449**	0.0689027
	TAP1	Chr. 6	**0.0003276**	**0.0025073**	**0.0118823**
	TRIM10	Chr. 6	**0.0000009**	**0.0002722**	**0.0000589**
	HLA-DRB1	Chr. 6	**0.0000018**	**0.0004108**	**0.0003191**
	LST1	Chr. 6	**0.0000009**	**0.0002963**	**0.0000931**
	DDX39B	Chr. 6	**0.0000008**	**0.0001398**	**0.0000000**
	MSH5	Chr. 6	**0.0005163**	**0.0009750**	**0.0154954**
	HLA-A	Chr. 6	**0.0016827**	**0.0046854**	**0.0318339**
Asthma Unrelated Genes	*PHIP*	Chr. 6	**0.0418319**	0.0986147	0.2299088
	EED	Chr. 11	**0.0094608**	**0.0100037**	0.0915084
	PYGB	Chr. 20	**0.0124950**	0.0805022	0.1086624
	SMARCD2	Chr. 17	**0.0131833**	**0.0153122**	0.1123372
	BORCS8	Chr. 19	**0.0233820**	0.1422346	0.1601654

$N_{case} = 46,802$ and $N_{control} = 347,481$; and Demenais et al. [29] with $N_{case} = 19,954$ and $N_{control} = 107,715$. We analyzed 15 genes, including 10 asthma-associated genes (*HLA-G*, *ATP6V1G2*, *TAP1*, *TRIM10*, *HLA-DRB1*, *LST1*, *HLA-DRB5*, *DDX39B*, *MSH5*, and *HLA-A*) in the MHC region and five genes that they are located outside the MHC region and no studies have reported association of asthma with the genes (*PHIP*, *EED*, *PYGB*, *SMARCD2*, and *BORCS8*) (Table S1).

We applied Fisher's (p_F) and our (p_N) methods to identify differentially expressed genes in this *meta*-TWAS. After applying Bonferroni's adjustment for multiple testing correction, the adjusted p-values are provided (Table 1). Fisher's method (p_F) identified all the ten asthma-associated genes, but also identified a high proportion of false-positive genes. Our method (p_N) identified most of the asthma-associated genes, except for *HLA-DRB5* (adjusted p-value = 0.069) and controlled false positive well. The results suggest that our method performs well and better than Fisher's method in combining more than two p-values.

Furthermore, our method can assign different weights to p-values in different TWASs. Here, the sample size of a study relative to the total sample size in the four TWASs was calculated as a weight; that is, a higher weight was assigned to a study with a higher sample size. Our weighted method (p_N^w) can identify all the ten asthma-associated genes, including *HLA-DRB5* that cannot be detected by the equal-weighted method (p_N). However, p_N^w obtained some false-positive findings, such as *EED* and *SMARCD2* in this analysis.

4. Conclusion and discussion

In this study, we proposed a novel numerical integration method (p_N) to evaluate statistical significance by combining correlated p-values from multiple studies in a *meta*-analysis. The proposed method is simple in concept and flexible to various correlation structures. Our theoretical investigation and simulation studies demonstrated that our proposed method performs well in terms of Type I error, statistical power, computing efficiency (regardless of the sample size), and statistical significance estimation accuracy. Real applications in large-scale GWASs and TWASs for rheumatoid arthritis and asthma facilitated efficient identification of critical genomic regions, such as the MHC region, and genes associated with rheumatoid arthritis and asthma, not reported in the previous GWASs or TWASs. We developed Pbine (https://github.com/Yinchun-Lin/Pbine) for *meta*-analysis based on a combination of p-values from multiple studies. Our method can com-

bine p-values from more than two studies with different sample sizes and precision so as to different levels of importance and information. The incorporation of unequal weights into a p-value combination for more than two studies has been implemented into Pbine.

The results in our simulation studies and real data analyses demonstrated that our method outperforms Fisher's method. We discussed the weakness of the decorrelation method – an order-noninterchangeable property. Although p_D can be calculated in either an ascending order or a descending order of p-values, we showed that p-values of these two procedures in the decorrelation method violate the uniformity property under a null distribution – the ascending order method (p_{DA}) tends to have more false negative and the descending order method (p_{DD}) have more false positive, particularly at the case with a high between-study correlation of p-values (Fig. S3). When we applied p_{DA} and p_{DD} in the *meta*-GWAS and *meta*-TWAS for rheumatoid arthritis, we did find a number of false-positive and false-negative findings (Table S2).

In addition to the methods discussed in this paper, studies have reported other p-value combination methods [30]. Some of these methods depend on p-value independency assumptions [31,32], parametric assumptions [33,34], and mathematical approximations such as Satterthwaite's approximation [35,36]. These methods may be efficient in computation. However, when their assumptions are violated, their performance is negatively affected by inflated Type I error, particularly when significance level α is low [37]. In addition, several methods have been developed on the basis of a generalization of Fisher's product p-value method, such as the weighted [38], truncated [15], and rank-truncated [39] product p-value methods. Our method can be generalized to more complicated cases.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.06.055.

References

[1] Fisher RA. Statistical methods for research workers. 4th ed. Edinburgh: Oliver and Boyd; 1932.

[2] Haidich AB. Meta-analysis in medical research. Hippokratia 2010;14(1):29–37.

[3] Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 2013;14(6):379–89.

[4] Hagg S et al. Gene-based meta-analysis of genome-wide association studies implicates new loci involved in obesity. Hum Mol Genet 2015;24(23):6849–60.

[5] Springelkamp H et al. Meta-analysis of genome-wide association studies identifies novel loci associated with optic disc morphology. Genet Epidemiol 2015;39(3):207–16.

[6] Scott EN et al. Transcriptome-wide association study uncovers the role of essential genes in anthracycline-induced cardiotoxicity. NPJ Genom Med 2021;6(1):35.

[7] Zeng P et al. Aggregating multiple expression prediction models improves the power of transcriptome-wide association studies. Hum Mol Genet 2021;30(10):939–51.

[8] Hu Y et al. A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet 2019;51(3):568–76.

[9] Li QZ et al. Fisher's method of combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations. Biostatistics 2014;15(2):284–95.

[10] Karlsson A. Permutation, parametric, and bootstrap tests of hypotheses. J Royal Stat Soc Series a-Stat Soc 2006;169:171.

[11] Hayes B. Overview of statistical methods for genome-wide association studies (GWAS). Methods Mol Biol 2013;1019:149–69.

[12] Feng H et al. Multitrait transcriptome-wide association study (TWAS) tests. Genet Epidemiol 2021;45(6):563–76.

[13] MacArthur J et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 2017;45(D1):D896–901.

[14] Beck T et al. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. Eur J Hum Genet 2014;22(7):949–52.

[15] Zaykin DV et al. Truncated product method for combining P-values. Genet Epidemiol 2002;22(2):170–85.

[16] Plenge RM et al. TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. N Engl J Med 2007;357(12):1199–209.

[17] Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661–78.

[18] Bonferroni, C.E., Teoria statistica delle classi e calcolo delle probabilit 'a. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936. 8.

[19] Newton JL et al. A review of the MHC genetics of rheumatoid arthritis. Genes Immun 2004;5(3):151–7.

[20] Matzaraki V et al. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol 2017;18(1):76.

[21] Zheng W, Rao S. Knowledge-based analysis of genetic associations of rheumatoid arthritis to inform studies searching for pleiotropic genes: a literature review and network analysis. Arthritis Res Ther 2015;17:202.

[22] Eyre S et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2012;44(12):1336–40.

[23] Stahl EA et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 2010;42(6):508–14.

[24] Cao C et al. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic Acids Res 2022;50(D1):D1123–30.

[25] Barbeira AN et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 2018;9(1):1825.

[26] Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. Nat Genet 2018;50(11):1593–9.

[27] Zhu Z et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. Nat Genet 2018;50(6):857–64.

[28] Zhu Z et al. Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. Eur Respir J 2019;54(6).

[29] Demenais F et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet 2018;50(1):42–53.

[30] Chen CW, Yang HC. OPATs: Omnibus P-value association tests. Brief Bioinform 2019;20(1):1–14.

[31] Lipták T. On the combination of independent tests.. A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményi 1958;3:171–97.

[32] Leroy Folks J. 6 Combination of independent tests. In: Handbook of Statistics. Elsevier; 1984. p. 113–21.

[33] Brown MB. 400: A method for combining non-independent, one-sided tests of significance. Biometrics 1975;31(4):987–92.

[34] Kost JT, McDermott MP. Combining dependent P-values. Stat Probab Lett 2002;60(2):183–90.

[35] Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics 1946;2(6):110–4.

[36] Dai H, Leeder JS, Cui Y. A modified generalized Fisher method for combining probabilities from dependent tests. Front Genet 2014;5:32.

[37] Zhang H, Wu Z. The generalized Fisher's combination and accurate p-value calculation under dependence. Biometrics 2022.

[38] Good I. On the weighted combination of significance tests. J Roy Stat Soc: Ser B (Methodol) 1955;17:264–5.

[39] Dudbridge F, Koeleman BP. Rank truncated product of P-values, with application to genomewide association scans. Genet Epidemiol 2003;25(4):360–6.