Received: 6 November 2017 | Revised: 4 March 2018 | Accepted: 16 March 2018

DOI: 10.1002/gepi.22123

RESEARCH ARTICLE

WILEY Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

Improved score statistics for meta-analysis in single-variant and gene-level association studies

Jingjing Yang[1,2] | Sai Chen[1] | Gonçalo Abecasis[1] | IAMDGC

[1]Department of Biostatistics, Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, United States of America

[2]Department of Human Genetics, Center for Computational and Quantitative Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America

Correspondence

Jingjing Yang, Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America.
Email: jyang51@emory.edu

Gonçalo Abecasis, Department of Biostatistics, Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, United States of America.
Email: goncalo@umich.edu

ABSTRACT

Meta-analysis is now an essential tool for genetic association studies, allowing them to combine large studies and greatly accelerating the pace of genetic discovery. Although the standard meta-analysis methods perform equivalently as the more cumbersome joint analysis under ideal settings, they result in substantial power loss under unbalanced settings with various case–control ratios. Here, we investigate the power loss problem by the standard meta-analysis methods for unbalanced studies, and further propose novel meta-analysis methods performing equivalently to the joint analysis under both balanced and unbalanced settings. We derive improved meta-score-statistics that can accurately approximate the joint-score-statistics with combined individual-level data, for both linear and logistic regression models, with and without covariates. In addition, we propose a novel approach to adjust for population stratification by correcting for known population structures through minor allele frequencies. In the simulated gene-level association studies under unbalanced settings, our method recovered up to 85% power loss caused by the standard methods. We further showed the power gain of our methods in gene-level tests with 26 unbalanced studies of age-related macular degeneration . In addition, we took the meta-analysis of three unbalanced studies of type 2 diabetes as an example to discuss the challenges of meta-analyzing multi-ethnic samples. In summary, our improved meta-score-statistics with corrections for population stratification can be used to construct both single-variant and gene-level association studies, providing a useful framework for ensuring well-powered, convenient, cross-study analyses.

KEYWORDS

meta-analysis, multi-ethnic studies, population stratification, score statistics, unbalanced studies

1 | INTRODUCTION

Meta-analysis is now an essential tool for genetic association studies, allowing them to combine information on 100,000s to 1,000,000s of samples, and greatly accelerating the pace of genetic discovery. Under ideal experiment settings, e.g., the same case–control ratio for all individual studies, the standard meta-analysis methods perform as efficiently as the more cumbersome alternative of joint analysis sharing individual-level data (Lin & Zeng, 2010). Standard meta-analysis methods have been routinely used in many large-scale genome-wide association studies (GWASs), identifying hundreds of complex trait loci, e.g., type 2 diabetes (T2D) (Fuchsberger et al., 2016; Scott et al., 2007; Zeggini et al., 2008), lipid levels (Willer et al., 2008), body mass index (BMI) (Willer et al., 2009), rheumatoid arthritis (Stahl et al., 2010), and fasting glucose levels (Prokopenko et al., 2009). Many tools implementing standard meta-analysis methods have been proposed for both single-variant and gene-level association studies, such as METAL for single-variant association studies (Willer, Li, & Abecasis, 2010), META-SKAT for sequential kernel association test (SKAT), MASS, and RAREMETAL for gene-level association studies (Feng, Liu, Zhan, Wing, & Abecasis, 2014; Lee, Teslovich, Boehnke, & Lin, 2013; Liu et al., 2014; Tang & Lin, 2013).

The standard meta-analysis methods generally sum the within-study test statistics as the meta-test-statistic that essentially eliminates all between-study variations, e.g., summing *P* values with respect to sample sizes (Stouffer et al., 1949), regression coefficients with respect to standard errors (Cochran, 1954), and score statistics with respect to variations (Lee et al., 2013). However, when the case–control ratios (or means and variances for quantitative traits) vary among individual studies due to unbalanced study designs, a common scenario for using Biobank data (Sudlow et al., 2015), the between-study variations due to the differences of case–control ratios actually contain important association information. This is why the standard meta-analysis methods ignoring between-study variations will lose power for unbalanced studies, compared to the joint analysis. Although the commonly used weighting strategy with respect to effective sample sizes may improve the power of the standard meta-analysis methods for single-variant association studies (Willer et al., 2010), it will fail for gene-level association studies based on score statistics such as META-SKAT (Lee et al., 2013) and RAREMETAL (Feng et al., 2014; Liu et al., 2014). This is because the magnitudes of score statistics are of the order of sample sizes (unlike the unit-free Z-score statistics in single-variant association studies).

Here, we describe a novel meta-analysis approach that models the between-study variances with improved meta-score-statistics, improving the power over the standard method under unbalanced settings. Our approach is suitable for both linear and logistic regression models, with and without covariates. When the study samples are of the same population (i.e., without population stratification), our meta-analysis methods are equivalent to the more cumbersome joint analyses (i.e., golden standards). For studies with multi-ethnic samples where the joint analysis is likely to cause inflated false positives, our methods will innovatively adjust for the population stratification using known population-specific minor allele frequencies (MAFs). Specifically, observing that the population stratification is reflected by different within-study MAFs in the score statistics, we will regress out the effects of known population-specific MAFs from the within-study MAFs. The population-specific MAFs are obtainable from reference panels such as 1000 Genome (Genomes Project et al., 2012), Biobanks (Sudlow et al., 2015), and gnomAD (Lek et al., 2016). In this paper, we focus on the meta-analysis methods with single-variant score test (Rao, 1948), gene-level Burden (Morris & Zeggini, 2010; Neale et al., 2011) test, and SKAT (Wu et al., 2011).

Our simulation studies showed that, under unbalanced settings, our methods recovered up to 84% power loss caused by the standard methods while controlling for false positive rates (i.e., type I errors), regardless of the existence of population stratification. Further, we demonstrated the power gain of our methods in the real gene-level association studies of age-related macular degeneration (AMD) (Fritsche et al., 2016), consisted with 26 unbalanced individual studies and 33,976 unrelated European samples (Table S1). For example, the known AMD risk gene CFI has SKAT *P* value 1.9 × 10^(−10) by joint analysis, *P* value 1.2 × [10]^(−4) by the standard meta-SKAT, and *P* value 3.1 × [10]^(−9) by our meta-SKAT. In addition, we applied our methods on the meta-analysis of three studies of T2D with Finnish and American European populations.

In summary, we propose novel meta-analysis methods based on our improved meta-score-statistics to achieve equivalent performance as joint analysis under unbalanced settings, for both single-variant and gene-level association studies. Our approach provides a useful framework for ensuring well-powered, convenient, cross-study analyses and is now implemented in the freely available RAREMETAL software.

2 | MATERIALS AND METHODS

2.1 | Score statistics for individual studies

Consider meta-analysis of K studies with n_k samples and m_k genotyped variants for the kth study. Let $\boldsymbol{y}_k$ denote the $n_k \times 1$ phenotype vector; $\boldsymbol{X}_k$ denote the $n_k \times m_k$ genotype matrix, encoding the minor allele count per individual per variant as (0, 1, 2); and $\boldsymbol{C_k}$ denote the $n_k \times (q_k + 1)$ augmented covariate matrix with the first column set to $1'$s and the others encoding q_k covariates. For each individual study, we consider the standard linear regression model (Equation 1) for quantitative traits

$$\boldsymbol{y_{ki}} = \boldsymbol{C_{ki}\alpha_k} + \boldsymbol{X_{ki}\beta_k} + \epsilon_i,\ \ \epsilon_i \sim \boldsymbol{N}\left(0,\ \boldsymbol{\sigma_k^2}\right),$$
$$\boldsymbol{i} = 1, \ldots,\ \boldsymbol{n_k}\ , \tag{1}$$

and the standard logistic regression model (Equation 2) for dichotomous traits

$$\boldsymbol{logit}\left(\boldsymbol{Prob}\left(\boldsymbol{y_{ki}} = 1\right)\right) = \boldsymbol{C_{ki}\alpha_k} + \boldsymbol{X_{ki}\beta_k}, \tag{2}$$

where $\boldsymbol{X}_{ki}$ is the ith row of genotype matrix $\boldsymbol{X}_k$, $\boldsymbol{\beta}_k$ is the vector of genetic effect sizes, $\boldsymbol{C}_{ki}$ is the ith row of augmented covariate matrix $\boldsymbol{C}_k$, and α_k is the vector of covariate effects including the intercept term. Let $\boldsymbol{u}_k$ denote the vector of score statistics for the kth study and $\boldsymbol{V}_k$ denote the variance–covariance matrix of $\boldsymbol{u}_k$ (Supplementary Appendix A).

2.2 | Standard meta-analysis

For notation simplicity, we assume the same set of variants for all K studies. The standard meta-analysis methods based on score statistics typically approximate the joint-score-statistics (obtainable in joint analysis) by

$$\boldsymbol{u} = \sum_{k=1}^{K} \boldsymbol{u}_k, \qquad \boldsymbol{V} = \sum_{k=1}^{K} \boldsymbol{V}_k. \tag{3}$$

Under unbalanced studies, these statistics will be systematically different from the joint-score-statistics, potentially leading to substantial power loss. Instead, we derive our improved meta-score-statistics $(\boldsymbol{u}, \boldsymbol{V})$ with summary-level data from the joint-score-statistic formulas with combined individual-level data.

2.3 | Simplified case without covariates

We first consider a simplified case without covariates, in which the following analytical formulas are derived from the joint-score-statistics under both linear and logistic regression models (Supplementary Appendix B.1), in terms of summary-level data including the within-study score statistics $(\boldsymbol{u}_k, \boldsymbol{V}_k)$, sample size n_k, phenotype mean deviation δ_k, residual variance estimate $\widehat{\sigma_k^2}$, and MAF vector $\boldsymbol{f}_k$

$$\boldsymbol{u} = \sum_{k=1}^{K} \boldsymbol{u}_k + \sum_{k=1}^{K} 2\boldsymbol{n}_k \boldsymbol{\delta}_k \left(\boldsymbol{f} - \boldsymbol{f}_k\right), \tag{4}$$

$$\boldsymbol{V} = \widetilde{\boldsymbol{\sigma}^2}\left[\sum_{k=1}^{K}\left[\frac{\boldsymbol{V}_k}{\widehat{\boldsymbol{\sigma}_k^2}}\right] - \sum_{k=1}^{K} 4\boldsymbol{n}_k\left(\boldsymbol{f}\boldsymbol{f}' - \boldsymbol{f}_k\boldsymbol{f}_k'\right)\right]. \tag{5}$$

Here, $\delta_k = (\frac{1}{n}\sum_{k=1}^{K} n_k\overline{y_k}) \ - \overline{y_k}$ denotes the difference between the overall phenotype mean and within-study phenotype mean; $\widetilde{\sigma^2} = \frac{1}{n-1}\ \sum_{k\,=\,1}^{K}[(n_k-1)\widehat{\sigma_k^2} + n_k\delta_k^2]$ denotes the joint residual variance; and $\boldsymbol{f} =$ denotes the overall MAF vector. The key difference from the standard approach is that we now model the between-study variations through the differences between the overall phenotype means, phenotype variances, and MAFs and their respective within-study values, as shown in the second term of Equations (4) and (5).

We note that, when $\delta_k = \ 0$, $\widehat{\sigma_k^2} \approx \widetilde{\sigma^2}$, $\boldsymbol{f}_k \approx \boldsymbol{f}$ as under balanced settings, both our meta-score-statistics (Equations 4 and 5) and the standard ones (Equation 3) are equivalent to the joint-score-statistics, which is why both methods perform as efficiently as the joint analysis for balanced studies. However, when $\delta_k \neq 0$, $\widehat{\sigma_k^2} \neq \widetilde{\sigma^2}$, $\boldsymbol{f}_k \neq \boldsymbol{f}$ as under unbalanced settings, the standard methods can no longer accurately approximate the joint-score-statistics, potentially leading to substantial power loss. In contrast, our meta-analysis methods will still be equivalent to the joint analysis.

2.4 | General case with covariates

Next, we consider the general case with covariates, in which our meta-score-statistic $\boldsymbol{u}$ is still derived as Equation (4) from the joint-score-statistic but our meta estimate of the joint variance–covariance matrix $\boldsymbol{V}$ will be different. For notation simplicity, we assume all individual studies have the same set of covariates. We approximate the phenotype mean deviation by $\delta_k \approx (\frac{1}{n}\sum_{k\,=\,1}^{K}(n_k\overline{\widehat{\mu_k}})) - \ \overline{\widehat{\mu_k}}$, where $\overline{\widehat{\mu_k}} = \frac{1}{n_k}\ \sum_{i\,=\,1}^{n_k}\widehat{\mu_{ki}}$ is the average of the fitted phenotypes in study k under the null regression models with $\boldsymbol{\beta} \ = \ 0$ (Equations 1 and 2).

Then under the linear regression model (Equation 1), we estimate $\boldsymbol{V}$ by

$$\begin{aligned}\boldsymbol{V} \approx \widetilde{\boldsymbol{\sigma}^2}\Bigg(&\sum_{k=1}^{K}\frac{\boldsymbol{V}_k}{\widehat{\boldsymbol{\sigma}_k^2}} + \sum_{k=1}^{K}\left(\boldsymbol{X}_k'\boldsymbol{C}_k\left(\boldsymbol{C}_k'\boldsymbol{C}_k\right)^{-1}\boldsymbol{C}_k'\boldsymbol{X}_k\right)\\ &-\left(\sum_{k=1}^{K}\boldsymbol{X}_k'\boldsymbol{C}_k\right)\left(\sum_{k=1}^{K}\boldsymbol{C}_k'\boldsymbol{C}_k\right)^{-1}\left(\sum_{k=1}^{K}\boldsymbol{X}_k'\boldsymbol{C}_k\right)'\Bigg),\end{aligned} \tag{6}$$

where the quantities of the covariate relationship matrix $\boldsymbol{C}_k'\boldsymbol{C}_k$ and genotype–covariate relationship matrix $\boldsymbol{X}_k'\boldsymbol{C}_k$ need to be shared across individual studies (see detailed derivations in Supplementary Appendix B.2).

Under the logistic regression model (Equation 2), we estimate $\boldsymbol{V}$ by

$$\begin{aligned}\boldsymbol{V} \approx &\sum_{k=1}^{K}\boldsymbol{V}_k + \sum_{k=1}^{K}\Delta_k\boldsymbol{X}_k'\boldsymbol{X}_k + \sum_{k=1}^{K}\left(\boldsymbol{X}_k'\widehat{\boldsymbol{P}_k}\boldsymbol{C}_k\right)\left(\boldsymbol{C}_k'\widehat{\boldsymbol{P}_k}\boldsymbol{C}_k\right)^{-1}\\ &\left(\boldsymbol{X}_k'\widehat{\boldsymbol{P}_k}\boldsymbol{C}_k\right)' - \left(\sum_{k=1}^{K}\left(\boldsymbol{X}_k'\widehat{\boldsymbol{P}_k}\boldsymbol{C}_k + \Delta_k\boldsymbol{X}_k'\boldsymbol{C}_k\right)\right)\\ &\left(\sum_{k=1}^{K}\left(\boldsymbol{C}_k'\widehat{\boldsymbol{P}_k}\boldsymbol{C}_k + \Delta_k\boldsymbol{C}_k'\boldsymbol{C}_k\right)\right)^{-1}\\ &\left(\sum_{k=1}^{K}\left(\boldsymbol{X}_k'\widehat{\boldsymbol{P}_k}\boldsymbol{C}_k + \Delta_k\boldsymbol{X}_k'\boldsymbol{C}_k\right)\right)',\end{aligned} \tag{7}$$

where $\widehat{\boldsymbol{P}_k} = \ diag(\widehat{\mu_{k1}}(1-\widehat{\mu_{k1}}), \ldots, \ \widehat{\mu_{kn_k}}(1-\widehat{\mu_{kn_k}}))$ denotes the diagonal matrix of phenotypic variances after correcting for within-study covariates; $\Delta_k = \ \delta_k \ (1 - 2\overline{\widehat{\mu_k}} - \delta_k)$ is the average difference between $\widehat{\boldsymbol{P}_k}$ and an analogous estimate in joint analysis (see detailed derivations in Supplementary Appendix B.2). To enable the calculation by Equation (7), the quantities of the genotype relation matrix $\boldsymbol{X}_k'\boldsymbol{X}_k$, covariate relation matrices $(\boldsymbol{C}_k'\boldsymbol{C}_k, \ \boldsymbol{C}_k'\widehat{\boldsymbol{P}_k}\boldsymbol{C}_k)$, and the genotype–covariate relation matrices $(\boldsymbol{X}_k'\boldsymbol{C}_k, \ \boldsymbol{X}_k'\widehat{\boldsymbol{P}_k}\boldsymbol{C}_k)$ need to be shared.

2.5 | Adjusting for population stratification

With multi-ethnic studies, our meta-analysis methods based on the improved meta-score-statistics (Equations 4–7; equivalent to the joint-score-statistics) will cause inflated false positives as in joint analysis. Thus, we have to adjust our meta-score-statistics to control for false positives caused by population stratification. Particularly, we note that the population stratification is reflected by the differences between the within-study and joint MAFs in the score statistics, e.g., $(\boldsymbol{f} - \boldsymbol{f}_k)$ in Equation (4) and $(\boldsymbol{f}\boldsymbol{f}' - \boldsymbol{f}_k\boldsymbol{f}_k')$ in Equation (5). Therefore, we propose to normalize our within-study MAFs by regressing out the population effects that can be explained by known population-specific MAFs. For example, with known MAF vectors $\boldsymbol{f}_{\mathrm{EUR}}$, $\boldsymbol{f}_{\mathrm{AMR}}$, $\boldsymbol{f}_{\mathrm{AFR}}$, $\boldsymbol{f}_{\mathrm{SAS}}$, $\boldsymbol{f}_{\mathrm{EAS}}$ of genome-wide variants for European, American, African, South Asian, and East Asian populations in the 1000 Genome Project (Genomes Project et al., 2012), we first fit the following linear regression model per individual study:

Then, in Equations (4) and (5), we substitute $\boldsymbol{f}_k$ by the residuals $\widetilde{\boldsymbol{\xi}_k} = \boldsymbol{f}_k - \sum \widehat{\gamma_{\mathrm{pop}}} \boldsymbol{f}_{\mathrm{pop}}$, and set $\boldsymbol{f}$ as the weighted residual averages $\frac{\sum_{k=1}^{K} n_k \boldsymbol{\xi}_k}{\sum_{k=1}^{K} n_k}$. For variants absent from the reference panel or with fitted values falling outside of the 95% predictive intervals, we set the corresponding elements in vectors $\boldsymbol{f}_k$ and $\boldsymbol{f}$ as 0 such that the between-study variances related to these variants will not be modeled by our methods. Equivalently, in Equations (6) and 2017, we can normalize the genotype matrix by $\tilde{\boldsymbol{X}} = \boldsymbol{X} - 2(\sum \widehat{\gamma_{\mathrm{pop}}} \boldsymbol{f}_{\mathrm{pop}})\boldsymbol{J}'$ for variants in the reference panel and set the genotype matrix as 0 for variants with unknown population-specific MAFs or with outlier fitted values.

Generally, we expect >99% R^2 for the model adjusting for population stratification, which requires reference panel that matches the ancestry of the study samples to provide population-specific MAFs. We also suggest matching reference ancestries to the study ancestries by using principle components, especially for admixed samples.

2.6 | Practical approach

Although Equations (6) and (7) enable corrections for covariates, they are generally not applicable in practice for the difficulties of sharing the quantities of $\boldsymbol{X}_k'\boldsymbol{X}_k$, $(\boldsymbol{C}_k'\boldsymbol{C}_k, \boldsymbol{C}_k'\widehat{\boldsymbol{P}}_k\boldsymbol{C}_k)$, and $(\boldsymbol{X}_k'\boldsymbol{C}_k, \boldsymbol{X}_k'\widehat{\boldsymbol{P}}_k\boldsymbol{C}_k)$. Thus, for computational simplicity, we suggest using Equation (5) with phenotypes corrected for covariates within individual studies under the linear regression model (Equation 1), where the dichotomous traits could be treated as quantitative traits by coding cases as $1'$s and controls as $0'$s. The RAREMETAL software also implements this practical approach. Both approaches (Equations 6 and 7 vs. Equation 5) produced nearly the same association results in our simulations. For both quantitative and dichotomous studies in this paper, we first corrected phenotypes within studies, and then used meta-score-statistics given by Equations (4) and (5) for association studies (adjusting for possible population stratification).

When correcting phenotypes for additional covariates by regression within individual studies, our meta-analysis methods require including the intercepts in the corrected phenotypes to correctly model the between-study variations. Otherwise, the phenotype deviation δ_k's will all be $0'$s, and our meta-score-statistics (Equation 4) will equal to the standard ones (Equation 3). In addition, we require the phenotype deviation δ_k's contain no other artificial effects (e.g., batch effects, effects due to different metrics or different underlying distributions across studies for phenotypes), because the between-study variations due to artificial effects are likely to cause inflated false positives.

2.7 | Test statistics

Our meta-analysis methods are based on accurately approximating the joint-score-statistics $(\boldsymbol{u}, \boldsymbol{V})$, and properly adjusting for possible population stratification. In this paper, we focus on score test for single-variant association studies, as well as the Burden test (Morris & Zeggini, 2010) with statistic $Q_{\mathrm{Burden}} = \frac{(w'u)^2}{w'Vw}$ and SKAT (Lee et al., 2013) with statistic $Q_{\mathrm{SKAT}} = \boldsymbol{u}'\boldsymbol{W}^2\boldsymbol{u}$ for gene-level association studies. Specifically, $\boldsymbol{w}' = (w_1, \ldots, w_m)$ is the variant-specific weight vector, and $\boldsymbol{W} = \mathrm{diag}(w_1, \ldots, w_m)$ is the $m \times m$ diagonal matrix. For each variant, we take the weight as "capped" beta density value $w_j = CBeta(f_j; 0.5, 0.5)$ with the corresponding MAF f_j, to avoid assigning extremely large weights for extremely rare variants ($Beta(f_j; 0.5, 0.5) \to \infty$ as $f_j \to 0$). That is, with sample size n, we have $CBeta(f_j; 0.5, 0.5) = Beta(\frac{5}{2n}; 0.5, 0.5)$ if the minor allele count $2nf_j < 5$, otherwise $CBeta(f_j; 0.5, 0.5) = Beta(f_j; 0.5, 0.5)$ allowing equal variance contributions from all variants.

Under the null hypothesis $(H_0 : \boldsymbol{\beta} = 0)$, both single-variant score statistic $Q_{\mathrm{score}} = \frac{\boldsymbol{u}^2}{\boldsymbol{V}}$ and Q_{Burden} follow a chi-square distribution with one degree of freedom $(df = 1)$. Under the null hypothesis $E(\beta_j) = 0,\ Var(\beta_j) = w_j^2\tau,\ j = 1, \ldots, m;\ \tau = 0)$ for SKAT, Q_{SKAT} asymptotically follows a mixture of chi-square distributions, $\sum_{j=1}^{m} \lambda_j \chi^2_{j,df=1}$, where $(\chi^2_{j,df=1})$ are independent chi-square random variables with $df = 1$, and λ_j's are nonzero eigenvalues of the variant relationship matrix $\boldsymbol{\Phi} = \boldsymbol{W}\boldsymbol{V}\boldsymbol{W}$.

2.8 | Simulation studies

To evaluate the false positive rate (type I error) and power of our meta-analysis methods, we conducted simulation studies in various scenarios with balanced and unbalanced studies, quantitative and dichotomous traits, with and without population stratification (see details of the simulation setup in Supplementary Appendix C).

Briefly, we first simulated haplotypes of three populations (European (EUR), Asian (ASA), and African (AFR)) by COSI with the well-calibrated coalescent model (Schaffner et al., 2005). Then we sampled genotypes of 1×10^5 individuals per population with 339 variants, 96% of which have MAFs < 5%. Random risk regions of 100 variants were selected to simulate both quantitative and dichotomous phenotypes, respectively, according to the standard linear and logistic models. We simulated phenotypes under the null models ($\boldsymbol{\beta} = 0$) for evaluating the empirical type I error, and phenotypes with 50% causal variants in the risk regions for evaluating the power.

We considered meta-analysis with five individual studies and a total sample size 3,000 (Table S2), under combined scenarios of dichotomous or quantitative traits, balanced or unbalanced settings, common or uncommon covariates, single- or multi-ethnic samples. For the balanced scenarios, each dichotomous study has 300 cases and 300 controls, while each quantitative study has 600 samples. For unbalanced dichotomous studies, there are (60, 180, 300, 420, 540) cases and (540, 420, 300, 180, 60) controls, such that the individual studies have the same sample size but various cases–control ratios. Unbalanced quantitative studies have sample sizes (200, 400, 600, 800, 1,000). Two covariate scenarios were simulated: (i) common covariates for all studies; (ii) different covariates among studies.

For the case with single-ethnic samples (i.e., without population stratification), we compared our adjusted meta-analysis methods with the standard methods and joint analysis, where the results by joint analysis will serve as golden standards. For the case with multi-ethnic samples (i.e., with population stratification; with EUR samples in studies 1 and 3, ASA samples in studies 2 and 4, and AFR samples in study 5), we only considered balanced and unbalanced dichotomous studies with common covariates (Table S2). In this case, we corrected the population stratification using the population MAF vectors ($\boldsymbol{f}_{EUR}$, $\boldsymbol{f}_{ASA}$, $\boldsymbol{f}_{AFR}$) that were calculated from 1×10^5 samples of the respective population. We compared our methods with the standard methods modeling no between-study variations and joint analysis correcting for population stratification by using the first four principle components (PCs) as additional covariates (Price et al., 2006).

2.9 | AMD and T2D data

The study of AMD by the International AMD Genomics Consortium (IAMDGC) (Fritsche et al., 2016) consists of 26 individual studies with 33,976 European, 1,572 Asian, and 413 African unrelated samples. Variants were genotyped on a customized Exome-Chip and imputed against the 1000 Genome Project Phase I reference panel. Advanced AMD cases include both cases with choroidal neovascularization and cases with geographic atrophy (Fritsche et al., 2016; Yang, Fritsche, Zhou, & Abecasis, 2017).

Three GWASs of T2D were considered in this paper: the Finland-United States Investigation of NIDDM genetics (FUSION) study (Scott et al., 2007), METabolic Syndrome In Men (METSIM) study (Laakso et al., 2017), and Michigan Genomics Initiative (MGI) study. We analyzed 2,297 unrelated Finnish samples (1,142 cases vs. 1,155 controls) in FUSION, 3,340 unrelated Finnish samples (673 cases vs. 2,667 controls) in METSIM, and 16,495 unrelated European American samples (1,942 cases vs. 14,553 controls) in MGI.

For the association studies of both AMD and T2D, all participants gave informed consent and the University of Michigan IRB approved our analyses.

3 | RESULTS

3.1 | Empirical type I errors in simulation studies

We repeated 2.5×10^7 null simulations per scenario to obtain empirical type I errors with significance levels $\alpha = 10^{-2}$, 10^{-4}, 2.5×10^{-6}. In the scenarios without population stratification, we showed that both Burden test and SKAT—by our adjusted meta-analysis method, the standard method, and joint analysis—have type I errors less than the corresponding significance levels (Fig. 1A; Supplementary Figs. S2 and S3). In the scenarios with population stratification, we showed that both Burden test and SKAT by our adjusted method and the standard method still controlled well for type I errors, with respective genomic control factors $\lambda_{\mathrm{GC}} = 0.995, 0.9926$ for Burden tests and $\lambda_{GC} = 0.7832, 0.9708$ for SKAT (Fig. 1B; Supplementary Figs. S4 and S5C–F). The slightly inflated type I error for standard SKAT is due to the arbitrary choice of variant-specific weighs that overweight the rare variants. In contrast, the joint analysis with first four joint PCs as additional covariates caused huge inflation with $\lambda_{\mathrm{GC}} = 11.9013, 52.6977$, respectively, for Burden test and SKAT (see Quantile-Quantile (QQ) plots of –log10(P values) in Supplementary Fig. S5A,B). This demonstrated that the standard methods modeling no between-study variations were free of population stratification, and that our approach of adjusting for population stratification successfully corrected for inflated type I errors.

3.2 | Empirical power in simulation studies

For each scenario, we repeated 10,000 simulations to obtain the empirical power that is given by the proportion of simulations with P values < 2.5×10^{-6} (genome-wide significance level for gene-level association tests). Here, our goal is to compare the power of our adjusted meta-analysis method with the standard method and the joint analysis. Here, the power differences between Burden test and SKAT depend on the simulation settings.

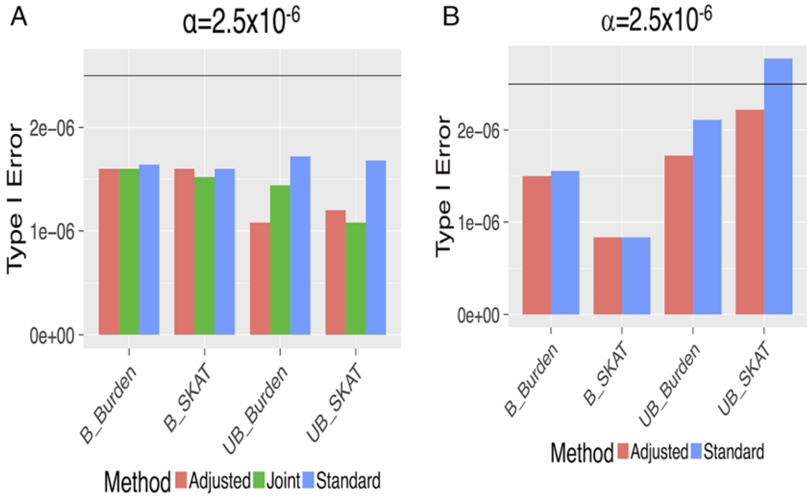

FIGURE 1 Empirical type I errors (with significance level $\alpha = 2.5 \times 10^{-6}$) for null simulations of balanced (B) and unbalanced (UB) dichotomous studies, with common covariates. (A) Scenario without population stratification. (B) Scenario with population stratification. “Adjusted” denotes our new meta-analysis methods; “Standard” denotes the standard meta-analysis methods; and “Joint” denotes the joint analyses using combined individual-level data

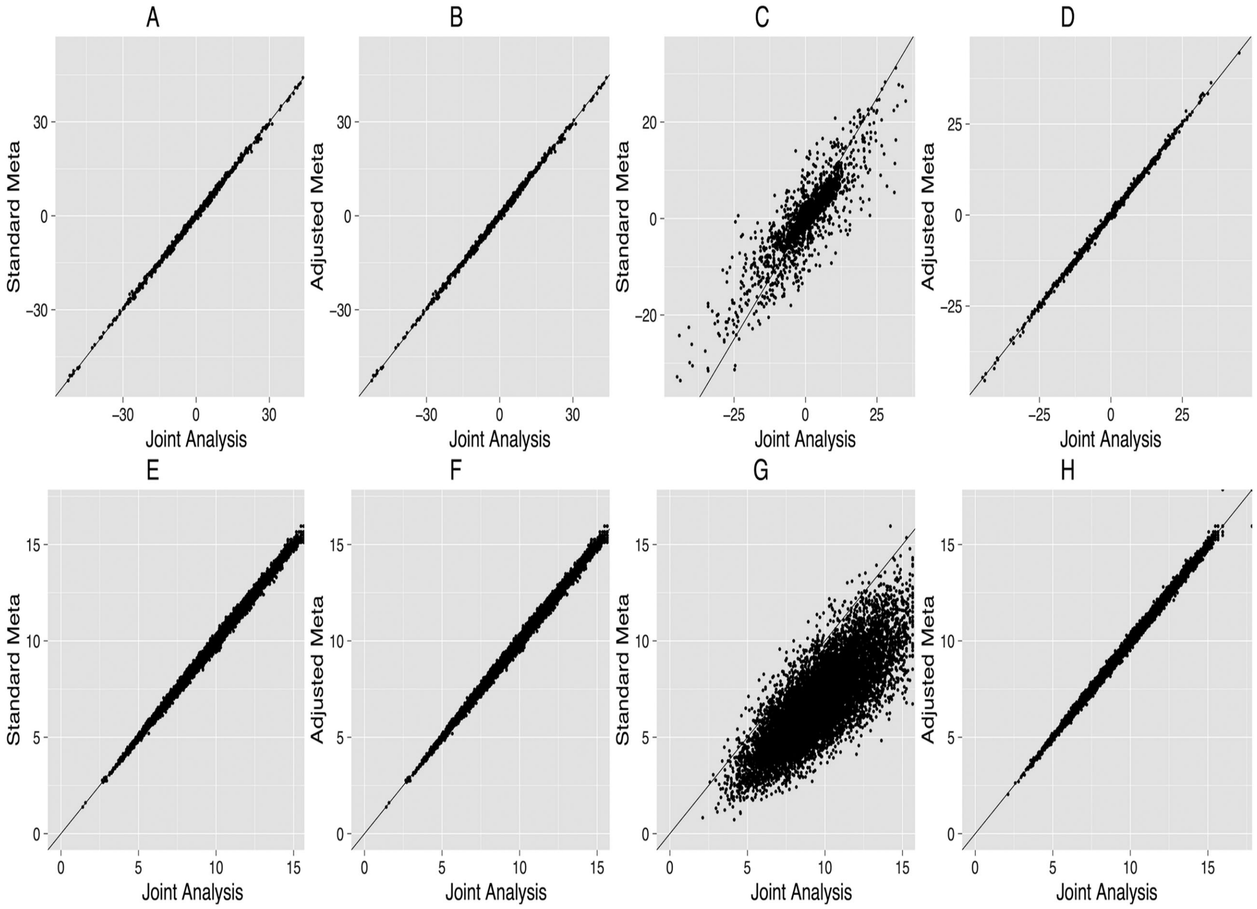

FIGURE 2 Score statistics (A, B, C, D) and –log10(*P* values) of the corresponding single-variant score tests (E, F, G, H), for dichotomous studies with common covariates, without population stratification, under balanced and unbalanced settings. (A, B) Score statistics under balanced studies. (C, D) Score statistics under unbalanced studies. (E, F) –log10(*P* values) of single-variant score tests under balanced studies. (G, H) –log10(*P* values) of single-variant score tests under unbalanced studies. “Standard Meta” denotes the standard meta-analysis methods; “Adjusted Meta” denotes our new meta-analysis methods

In the balanced dichotomous studies without population stratification, both standard and our adjusted meta-score-statistic estimates were highly concordant with the golden standards obtained by joint analysis ($R^2 > 99.8\%$; Fig. 2A, B). In the unbalanced dichotomous studies, the standard meta-score-statistic estimates scattered further from the joint-score-statistics ($R^2 \sim 78.2\%$, Fig. 2C), while our estimates were still concordant with the joint-score-statistics ($R2 > 99.8\%$; Fig. 2D). Consequently, under balanced settings, the *P* values of single-variant score tests by both standard methods and our adjusted methods were concordant with the joint analysis results (Fig. 2E, F). While under unbalanced settings, the *P* values by standard methods were less significant than joint analysis results (Fig. 2G), hence less significant than the results by our adjusted methods that were concordant with the joint analysis results (Fig. 2H).

In the scenarios without population stratification, the gene-level tests (i.e., Burden test and SKAT) based on our improved meta-score-statistic estimates are equivalent to the ones in joint analysis under general settings, recovering up to 69% power loss caused by the standard method in unbalanced dichotomous studies with common covariates (Fig. 3). Similar results were obtained for scenarios with different covariates (Supplementary Figs. S6–S8). Take the dichotomous studies with common covariates for example (Fig. 3), the power by standard meta-analysis method was 0.701 for Burden and 0.219 for SKAT, which were 27% and 69% less than the golden standards (0.964 for Burden; 0.703 for SKAT) by joint analysis; while the results by our adjusted meta-analysis method (power 0.964 for Burden; 0.702 for SKAT) were concordant with the joint analysis results.

In the scenarios with population stratification, the joint analysis (with top four PCs as additional covariates) no longer provide golden standards due to highly inflated type I errors with $\lambda_{GC} = 11.9013$, 52.6977 Burden test and SKAT, respectively (Fig. S5A, B). Hence, we only compared the empirical powers by our adjusted meta-analysis method with the standard method. Again, both methods had similar power in balanced dichotomous studies, while our adjusted meta-analysis method recovered up to 85% power loss by the standard method in unbalanced dichotomous studies (0.898 vs. 0.302 for Burden test, Fig. 3C; 0.880 vs. 0.126 for SKAT, Fig. 3D).

For quantitative studies, although we simulated "unbalanced" scenarios with various sample sizes, these are not really unbalanced for having similar phenotype means across individual studies (i.e., the between-study variances were close to 0). As a result, both our adjusted method and the standard method produced equivalent results as the joint analyses under all settings (Supplementary Figs. S9–S13).

In summary, the simulation studies showed that our adjusted meta-analysis method will improve power by correctly modeling the association information in the between-study variances. When the between-study variances are close to 0 as under balanced settings, both our method and the standard method are equivalent to the joint analysis. When the between-study variances are also subject to population stratification, our method require known population-specific MAFs to correct for possibly inflated type I errors.

3.3 | Real study of AMD

We applied our method on the real AMD data collected by the IAMDGC (Fritsche et al., 2016), which has 26 individual studies with 33,976 European, 1,572 Asian, and 413 African unrelated samples. We treated the Asian and African samples as two extra studies. First, we conducted null simulations for 2.5×10^7 times using the AMD data, by permuting the real AMD phenotypes and randomly selecting genotype regions of 100 variants for Burden test and SKAT. We found that both our adjusted and the standard meta-analysis methods had type I errors around the significance level, while the joint analyses with first 4 joint PCs as extra covariates produced clearly inflated type I errors (Supplementary Fig. S14). Specifically, with significance level 2.5×10^{-6}, the joint analyses (Joint_PC4) had type I errors 8.6×10^{-6} for Burden test and 9.2×10^{-6} for SKAT.

For valid comparisons with joint analyses, we only considered European samples from the 26 unbalanced studies (Table S1) for Burden test and SKAT in 3 example AMD risk genes (Fritsche et al., 2016) (*CFH*, *CFI*, *TIMP3*). Previous analyses by variable-threshold tests (Price et al., 2010) (with respective MAF thresholds 0.015%, 0.068%, 0.021% for genes *CFH*, *CFI*, *TIMP3*) gave significant p values ($< 2.5 \times 10^{-6}$) for these three loci. To be consistent with the previous variable-threshold tests (Price et al., 2010), we only analyzed protein-altering variants (imputed/genotyped) with MAFs under the corresponding thresholds (MAFs < 0.015%, 0.068%, 0.021%), and corrected for the same covariates—known independent signals within the same locus, gender, first two principal components (calculated using the combined data), and source of DNA (whole-blood or whole genome-amplified DNA).

Our adjusted meta-analysis method produced genome-wide significant p values for genes *CFH* and *CFI* (Table 1), which were more significant than the ones by the standard method. Specifically, gene *CFH* had genome-wide significant Burden *P* value 2.4×10^{-7} by joint analysis, versus 2.1×10^{-6} by our adjusted method and 3.2×10^{-5} by the standard method (no longer genome-wide significant). Although all methods obtained significant Burden *P* values for gene *CFI*, the *P* value by our method was still more significant than the one by the standard method (3.3×10^{-14} vs. 9.6×10^{-10}) and closer to the *P* value by joint analysis (8.9×10^{-15}). Similarly, the SKAT p value by standard method for gene *CFI* was no longer genome-wide significant (1.2×10^{-4}), while the

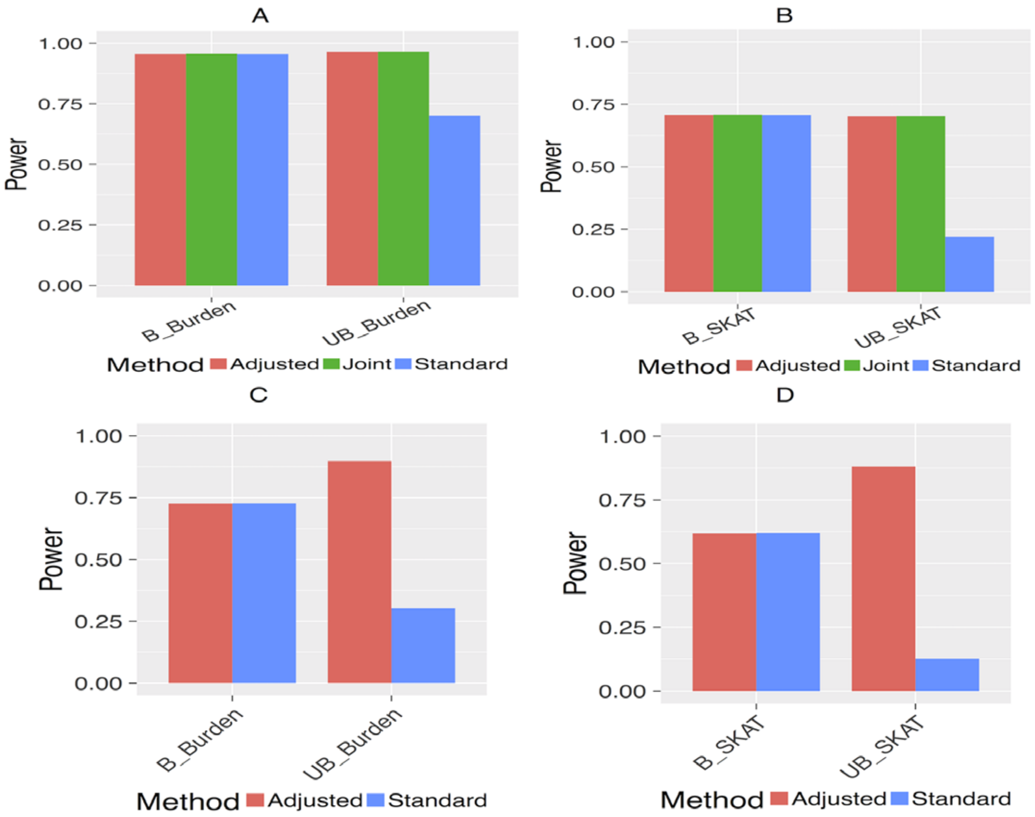

FIGURE 3 Power comparisons of meta-Burden test and meta-SKAT, for balanced (B) and unbalanced (UB) dichotomous studies with common covariates. (A, B) Without population stratification. (C, D) With population stratification. "Adjusted" denotes our new meta-analysis methods; "Standard" denotes the standard meta-analysis methods; and "Joint" denotes the joint analyses using combined individual-level data

TABLE 1 *P* values of gene-level Burden test and SKAT by the standard meta-analysis method, our adjusted meta-analysis method, and joint analysis

	Burden test			SKAT		
Gene	**Standard**	**Adjusted**	**Joint**	**Standard**	**Adjusted**	**Joint**
CFH	3.2×10^{-5}	2.1×10^{-6}	2.4×10^{-7}	6.1×10^{-4}	8.4×10^{-5}	3.6×10^{-5}
CFI	9.6×10^{-10}	3.3×10^{-14}	8.9×10^{-15}	1.2×10^{-4}	3.1×10^{-9}	1.9×10^{-10}
TIMP3	9.8×10^{-4}	1.0×10^{-5}	1.8×10^{-5}	2.6×10^{-3}	7.4×10^{-5}	2.6×10^{-4}

SKAT p value by our adjusted method (3.1×10^{-9}) was still genome-wide significant and close to the one by joint analysis (1.9×10^{-10}).

Even though all approaches failed to identify the *TIMP3* locus with *P* values 1.8×10^{-5} by joint Burden test and 2.6×10^{-4} by joint SKAT, our method still produced more significant *P* values than the standard method (1.0×10^{-5} vs. 9.8×10^{-4} for Burden test; 7.4×10^{-5} vs. 2.6×10^{-3} for SKAT). Likely due to the random errors between meta-score-statistic estimates and the joint-score-statistics, the Burden and SKAT p values for gene *TIMP3* by our method are slightly smaller than the ones by joint analysis.

This real example of AMD study demonstrated that our method produced similar results as the joint analysis (golden standards with single-ethnic samples), recovering the power loss by the standard method.

3.4 | Real study of T2D

In this real example, we considered single-variant meta-analyses of three T2D GWASs: FUSION (1,142 cases vs. 1,155 controls; unrelated Finnish samples) (Scott et al., 2007), METSIM (673 cases vs. 2,667 controls; unrelated Finnish male samples) (Laakso et al., 2017), and MGI (1,942 cases vs. 14,553 controls; unrelated European American samples). These three unbalanced GWASs have various case–control ratios (0.98, 0.24, 0.13) and multi-ethnic samples (Supplementary Figs. S15 and S16).

We first jointly corrected the T2D phenotypes for age, gender, BMI, and first two joint PCs, within individual studies. The reason of jointly correcting the T2D phenotypes is to eliminate the possible between-study variance due to the artificial effects caused by individually corrected phenotypes. Then we applied the joint analysis, the standard meta-analysis

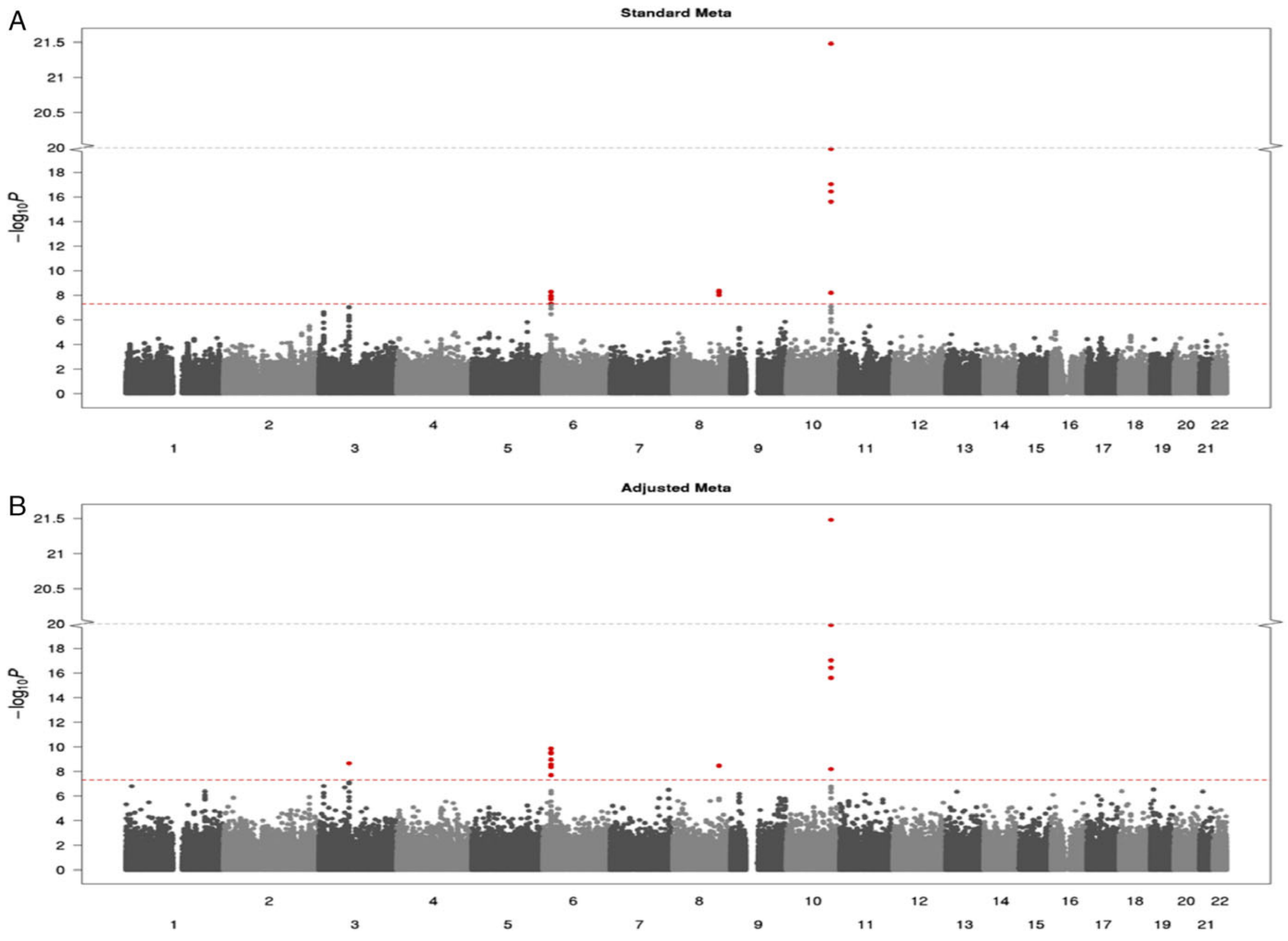

FIGURE 4 Manhattan plots of meta GWASs of type 2 diabetes, by (A) standard method and (B) our adjusted method. "Standard Meta" denotes the standard meta-analysis methods; "Adjusted Meta" denotes our new meta-analysis methods

method, our joint-equivalent meta-analysis method (without adjustment for population stratification), and our adjusted meta-analysis method with adjustment for population stratification using the population-specific MAFs of EUR, AMR, AFR, SAS, EAS from the 1000 Genome Project (Genomes Project et al., 2012) (~500 samples per population).

In this study, we only analyzed 631,870 variants genotyped in the METSIM study. These analyzed variants could be either genotyped or imputed to 1000 Genome Project (Genomes Project et al., 2012) or absent in FUSION (627,920 variants) and MGI (631,628 variants) studies (see Manhattan plots of the individual GWASs in Supplementary Fig. S17). As expected, the joint analysis and the joint-equivalent methods resulted in inflation with $\lambda_{GC} =$ 1.11, 1.13, while the standard meta-analysis method had $\lambda_{\mathrm{GC}} =$ 1.07 (Supplementary Fig. S18). Specifically, the standard method identified three known T2D risk loci (*SLC30A8* on CHR8, *TCF7L2* on CHR10, and *CDKAL1* on CHR6) (Billings & Florez, 2010), while our method with adjustment for population stratification identified comparable *P* values for signals in the *SLC30A8* and *TCF7L2* loci, more significant *P* value in the *CDKAL1* locus, and one extra potential loci *ROBO2* on CHR3 (see Manhattan plots in Fig. 4).

We looked into the within-study MAFs of all "genome-wide significant" variants that were identified by joint analysis (Supplementary Figs. S19 and S20). We noticed that the known signals generally have comparable MAFs across three studies, especially for MAFs between FUSION and METSIM with Finnish samples. The variants with large variation among within-study MAFs are likely to be "false positives," according to our improved meta-score-statistic formulas. Although our adjusted method corrected for these likely "false positives" signals, it failed to completely correct for the inflation with $\lambda_{\mathrm{GC}} =$ 1.15 (Supplementary Fig. S18D). This is probably because the 1000 Genome variants fail to provide a good reference MAFs for the Finnish samples in FUSION and METSIM studies. Specifically, the regression R^2s were 97.1%, 96.3%, and 99.5% for regressing known population-specific effects (MAFs) out from the within-study MAFs in the FUSION, METSIM, and MGI studies, respectively.

This real study demonstrated the benefit of improving power by applying our adjusted meta-analysis method on unbalanced studies. Further, this study showed the challenges of correctly adjusting for population stratification when samples are multi-ethnic, which requires good reference panels to provide population-specific MAFs.

4 | DISCUSSION

In this paper, we propose improved meta-score-statistics in terms of summary-level data that retain $R^2 > 99\%$ with the joint-score-statistics using individual-level data, under general settings. We derived our adjusted meta-analysis methods based on the improved meta-score-statistics for both single-variant and gene-level association studies, performing equivalently with the joint analysis. We further propose a novel approach to adjust for population stratification by using the known population-specific MAFs. By extensive simulation and real studies, we demonstrated that our adjusted meta-analysis methods controlled well for type I errors and gained power over the standard meta-analysis methods for unbalanced studies.

Although we derived the improved meta-score-statistics for both linear and logistic regression models, with and without covariates, we suggest using the simplified formula (Equation 5) that requires to share score statistics, sample sizes, means, and variances of phenotypes corrected for covariates, and MAFs, for practical usage. Alternatively, the more complicated formula (Equation 6 or Equation 7) requires to share additional genotype, genotype–covariate, and covariate relationship matrices for incorporating covariates, thus generally more challenging in practice. Different from the standard meta-analysis methods that only share score statistics, sharing the extra summary data enables our methods to model the between-study variations, thus performing as efficiently as the joint analysis.

Of course, our meta-analysis methods are not without limitations. First, our method assumes that the genetic effects are homogeneous across studies and the phenotypes are of the same distribution. Second, our method requires that there are no confounded artificial effects in the between-study variances, otherwise the standard methods will be preferred. Third, our method requires the population-specific MAFs from good reference panels to correctly adjust for population stratification, where we suggest locating external samples of the same ancestries as the individual studies by using principle components.

In conclusion, sharing summary data allows us to leverage the power of large sample sizes without the hassle of combining individual-level data, and then helps identify more genetic risk loci for complex traits. We illustrated that the standard meta-analysis methods will lose power under unbalanced studies for not modeling the association information in the between-study variations. Whereas our adjusted meta-analysis methods that correctly model the between-study variations will improve the power under unbalanced settings, providing a useful framework to ensure well-powered, convenient, cross-study association analyses.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIH) grant R01HG007022 and the start-up fund provided by Department of Human Genetics at the Emory University School of Medicine. The authors would like to thank Dr. Dajiang Liu (Pennsylvania State University), Dr. Shuang Feng (graduated Ph.D. student from University of Michigan), and Dr. Shawn Lee (University of Michigan) for their inspiration and valuable comments on this work. Especially, the authors want to thank all investigators of the IAMDGC, FUSION, METSIM, and MGI studies for providing the real data of the AMD and T2D studies in this work.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Jingjing Yang http://orcid.org/0000-0002-4191-4138

REFERENCES

Billings, L. K., & Florez, J. C. (2010). The genetics of type 2 diabetes: What have we learned from GWAS? *Annals of the New York Academy of Sciences*, *1212*, 59–77. https://doi.org/10.1111/j.1749-6632.2010.05838.x

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*(1), 101–129. https://doi.org/10.2307/3001666

Feng, S., Liu, D., Zhan, X., Wing, M. K., & Abecasis, G. R. (2014). RAREMETAL: Fast and powerful meta-analysis for rare variants. *Bioinformatics*, *30*(19), 2828–2829. https://doi.org/10.1093/bioinformatics/btu367

Fritsche, L. G., Igl, W., Bailey, J. N., Grassmann, F., Sengupta, S., Bragg-Gresham, J. L., … Heid, I. M. (2016). A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nature Genetics*, *48*(2), 134–143. https://doi.org/10.1038/ng.3448

Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., … McCarthy, M. I. (2016). The genetic architecture of type 2 diabetes. *Nature*, *536*(7614), 41–47. https://doi.org/10.1038/nature18642

Genomes Project Consortium, Abecasis, R., G., Auton, A., Brooks, L. D., DePristo, M. A., ... McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. https://doi.org/10.1038/nature11632

Laakso, M., Kuusisto, J., Stancakova, A., Kuulasmaa, T., Pajukanta, P., Lusis, A. J., … Boehnke, M. (2017). The metabolic syndrome in men study: A resource for studies of metabolic and cardiovascular diseases. *Journal of Lipid Research*, *58*(3), 481–493. https://doi.org/10.1194/jlr.O072629

Lee, S., Teslovich, T. M., Boehnke, M., & Lin, X. (2013). General framework for meta-analysis of rare variants in sequencing association studies. *American Journal of Human Genetics*, *93*(1), 42–53. https://doi.org/10.1016/j.ajhg.2013.05.010

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., & Fennell, T. (2016). … Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285–291. https://doi.org/10.1038/nature19057

Lin, D. Y., & Zeng, D. (2010). Meta-analysis of genome-wide association studies: No efficiency gain in using individual participant data. *Genetic Epidemiology*, *34*(1), 60–66. https://doi.org/10.1002/gepi.20435

Liu, D. J., Peloso, G. M., Zhan, X., Holmen, O. L., Zawistowski, M., Feng, S., … Abecasis, G. R. (2014). Meta-analysis of gene-level tests for rare variant association. *Nature Genetics*, *46*(2), 200–204. https://doi.org/10.1038/ng.2852

Morris, A. P., & Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genetic Epidemiology*, *34*(2), 188–193. https://doi.org/10.1002/gepi.20450

Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., … Daly, M. J. (2011). Testing for an unusual distribution of rare variants. *Plos Genetics*, *7*(3), e1001322. https://doi.org/10.1371/journal.pgen.1001322

Price, A. L., Kryukov, G. V., de Bakker, P. I., Purcell, S. M., Staples, J., Wei, L. J., & Sunyaev, S. R. (2010). Pooled association tests for rare variants in exon-resequencing studies. *American Journal of Human Genetics*, *86*(6), 832–838. https://doi.org/10.1016/j.ajhg.2010.04.005

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

Prokopenko, I., Langenberg, C., Florez, J. C., Saxena, R., Soranzo, N., Thorleifsson, G., … Abecasis, G. R. (2009). Variants in MTNR1B influence fasting glucose levels. *Nature Genetics*, *41*(1), 77–81. https://doi.org/10.1038/ng.290

Rao, R. C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, *44*(01), 50–57.

Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J., & Altshuler, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. *Genome Research*, *15*(11), 1576–1583. https://doi.org/10.1101/gr.3709305

Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., … Boehnke, M. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, *316*(5829), 1341–1345. https://doi.org/10.1126/science.1142382

Stahl, E. A., Raychaudhuri, S., Remmers, E. F., Xie, G., Eyre, S., Thomson, B. P., … Plenge, R. M. (2010). Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genetics*, *42*(6), 508–514. https://doi.org/10.1038/ng.582

Stouffer, S. A., Schuman, E. A., DeVinney, L. C., Star, S., & Williams, R. M. (1949). *The American Soldier, Adjustment during army life* (Vol. 1). Oxford, England: Princeton University Press.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *Plos Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Tang, Z. Z., & Lin, D. Y. (2013). MASS: Meta-analysis of score statistics for sequencing studies. *Bioinformatics*, *29*(14), 1803–1805. https://doi.org/10.1093/bioinformatics/btt280

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., … Abecasis, G. R. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature Genetics*, *40*(2), 161–169. https://doi.org/10.1038/ng.76

Willer, C. J., Speliotes, E. K., Loos, R. J., Li, S., Lindgren, C. M., & Heid, I. M.,. … Genetic Investigation of Anthropometric Traits Consortium. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics*, *41*(1), 25–34. https://doi.org/10.1038/ng.287

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*, *89*(1), 82–93. https://doi.org/10.1016/j.ajhg.2011.05.029

Yang, J., Fritsche, L. G., Zhou, X., & Abecasis, G., & International Age-Related Macular Degeneration Genomics Consortium. (2017). A scalable bayesian method for integrating functional information in genome-wide association studies. *American Journal of Human Genetics*, *101*(3):404–416. https://doi.org/10.1016/j.ajhg.2017.08.002

Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., … Altshuler, D. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics*, *40*(5), 638–645. https://doi.org/10.1038/ng.120

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Yang J, Chen S, Abecasis G, IAMDGC. Improved score statistics for meta-analysis in single-variant and gene-level association studies. *Genetic Epidemiology*. 2018;42:333–343. https://doi.org/10.1002/gepi.22123