

Meta-Analysis of Genetic Association Studies and Adjustment for Multiple Testing of Correlated SNPs and Traits

Karen N. Conneely^{1*} and Michael Boehnke²

¹Department of Human Genetics, Emory University, Atlanta, Georgia

²Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan

Meta-analysis has become a key component of well-designed genetic association studies due to the boost in statistical power achieved by combining results across multiple samples of individuals and the need to validate observed associations in independent studies. Meta-analyses of genetic association studies based on multiple SNPs and traits are subject to the same multiple testing issues as single-sample studies, but it is often difficult to adjust accurately for the multiple tests. Procedures such as Bonferroni may control the type-I error rate but will generally provide an overly harsh correction if SNPs or traits are correlated. Depending on study design, availability of individual-level data, and computational requirements, permutation testing may not be feasible in a meta-analysis framework. In this article, we present methods for adjusting for multiple correlated tests under several study designs commonly employed in meta-analyses of genetic association tests. Our methods are applicable to both prospective meta-analyses in which several samples of individuals are analyzed with the intent to combine results, and retrospective meta-analyses, in which results from published studies are combined, including situations in which (1) individual-level data are unavailable, and (2) different sets of SNPs are genotyped in different studies due to random missingness or two-stage design. We show through simulation that our methods accurately control the rate of type I error and achieve improved power over multiple testing adjustments that do not account for correlation between SNPs or traits. *Genet. Epidemiol.* 34:739–746, 2010. © 2010 Wiley-Liss, Inc.

Key words: meta-analysis; association study; multiple testing; SNPs

Contract grant sponsor: National Institutes of Health; Contract grant number: HG000376.

*Correspondence to: Karen N. Conneely, Department of Human Genetics, Emory University, 615 Michael Street Suite 301, Atlanta, GA 30322. E-mail: kconnee@emory.edu

Received 21 April 2010; Revised 15 July 2010; Accepted 18 August 2010

Published online 27 September 2010 in Wiley Online Library (wileyonlinelibrary.com).

DOI: 10.1002/gepi.20538

INTRODUCTION

The large sample sizes necessary to detect subtle genetic effects are often attainable only through the combination of samples of individuals across multiple studies. When significant genetic associations are detected, replication in additional studies is essential to validate initial results. For these reasons, meta-analysis is now standard in both candidate gene and genome-wide association studies (GWAS). This presents a unique multiple testing problem, since standard approaches to multiple testing correction that require access to individual-level data may not be useable in this case. For GWAS, the common approach of defining significance based on a pre-set genome-wide cutoff such as $P < 5 \times 10^{-8}$ can be easily applied to meta-analyses as well. However, candidate gene studies have study-specific multiple testing burdens requiring significance thresholds to be customized based on the number of tests and degree of correlation between the tests. In many candidate gene meta-analyses, computation of the appropriate significance level will be complicated by the unavailability of individual-level data.

Since the first meta-analysis was published by Pearson [1904], most meta-analyses have been retrospective studies carried out by combining summary statistics from previously

published studies. The lack of access to individual-level data for these studies necessitated the development of meta-analysis techniques to appropriately combine summary statistics, including the popular Mantel and Haenszel [1959] test for combining 2×2 tables and the inverse-variance weighting approach of Woolf [1955].

In contrast, meta-analyses in the current genetic association literature often employ a prospective study design in which the data are analyzed with the intent to combine results across studies; these prospective meta-analyses often take place as collaborations between multiple centers. When individual data can be combined for analysis, a meta-analysis can be performed in a regression context by including fixed or random effects to control for genetic and/or phenotypic differences between centers. However, due to limitations on data sharing, it is not always possible to combine individual-level data in a single analysis. Thus, in many cases, prospective consortia-based meta-analyses will face the same issues as retrospective meta-analyses, and will be forced to rely on meta-analysis techniques to combine summary statistics such as counts, test statistics, or P -values across studies. Although no efficiency is lost when meta-analyses are performed using summary statistics rather than individual-level data [Lin and Zeng, 2009], lack of access to individual-level data may limit the available options for multiple-testing adjustment.

Several methods of meta-analysis are commonly used in genetic association studies. In case-control studies or other studies involving binary (dichotomous) outcome variables, SNPs are often tested for association via the Cochran-Armitage test for trend [Cochran, 1954; Armitage, 1955]. To combine the results of several trend-test-based studies, a natural approach is the Cochran-Mantel-Haenszel test [Mantel, 1963], which extends the Mantel and Haenszel [1959] test for combining 2×2 tables to a trend test setting. For studies involving quantitative traits or more complicated statistical models, more general approaches for combining results such as the inverse-variance approach [Woolf, 1955] or sample size weighting can be used. These methods can be applied to tests with quantitative or binary response variables, and dose variables that are continuous rather than integer-valued, and are especially useful in genetic studies involving quantitative traits, environmental and demographic covariates, and continuous dose variables such as probabilistic imputed genotype scores [Li et al., 2009].

Although the methods typically used to combine data across genetic association studies are well-established, only recently have these methods been applied to the multitude of SNPs and traits tested in genetic association studies, so adjustment of meta test statistics for multiple testing has not been addressed. Bonferroni-type methods that do not account for correlation between tests usually will be conservative, given the degree of correlation between dense SNPs and between related traits. Permutation tests are generally not possible in retrospective meta-analyses, and also may not be possible in prospective meta-analyses if individual-level data cannot be shared across studies.

We previously described P_{ACT} , a multiple-testing adjustment that accounts for the correlation between tests and provides a faster alternative to permutation testing [Conneely and Boehnke, 2007]. P_{ACT} (P -value adjusted for correlated tests) can be used to adjust the most significant P -values or test statistics from tests of K traits for association with M genetic variants. We showed that in a generalized linear model framework, the $K \times M$ test statistics often follow an asymptotic multivariate normal distribution $N(0, \mathbf{R})$, where \mathbf{R} is the correlation matrix corresponding to the covariance matrix \mathbf{V} , the Kronecker product of the sample covariance matrices of traits and genotypes, conditioned on covariates. Here, we show that this result readily extends to a general class of meta-analyses that includes all those described above. We describe how P_{ACT} can be applied to meta-analyses with the prospective or retrospective designs discussed above, as well as to meta-analyses in which SNPs are not all genotyped in all studies.

Data can be missing on many levels in genetic studies. Within a single study, individuals may be missing data for certain traits or SNP genotypes. In meta-analyses involving multiple studies, certain SNPs or traits may not be available in all studies due to decisions made prior to data analysis. For example, SNPs may be genotyped in some studies but not others due to platform differences, assay failure, poor performance on quality control measures, or constrained resources. When missingness occurs independently of association results in other studies, it is straightforward to deal with in the context of computing meta-statistics and adjusting them for multiple testing with P_{ACT} .

In contrast, when the association results observed in an initial study determine which SNP or trait data are analyzed in subsequent studies, a different approach is

needed. In two-stage analyses, many tests may be performed on an initial sample of individuals, but only SNPs or traits passing a pre-set significance criterion in stage one are followed up in stage two. Skol et al. [2006, 2007] present a sample-size weighted meta test statistic that accounts for the conditional selection of SNPs for inclusion in stage two. Multiple testing adjustment of two-stage analyses cannot be performed with standard permutation tests, although alternative methods have been proposed [Lin, 2006; Dudbridge, 2006]. We show here how P_{ACT} can be used in two-stage studies to adjust for multiple testing while accounting for the correlation between tests.

Finally, we present simulations to assess the validity and power of our approach in the situations described above. Our simulations suggest that our method provides a valid adjustment for correlated meta-analysis statistics with Cochran-Mantel-Haenszel, inverse variance, and sample-size-weighted meta-statistics, with binary or quantitative traits, with prospective or retrospective design, and with SNPs missing for an entire sample of individuals either at random or through threshold-based selection of follow-up SNPs in two-stage analyses.

MATERIALS AND METHODS

META-ANALYSIS TECHNIQUES

The Cochran-Armitage trend test [Cochran, 1954; Armitage, 1955], a special case of the score test from a logistic regression, can be used to test for a linear relationship between an ordered “dose” variable (for example, a dose of a treatment) and the log odds of being a case vs. a control. This test is commonly used in genetic association studies to model additive genetic effects by defining “dose” as a genotype score defined as the number of copies of a reference allele (0, 1, or 2); this is the genotypic trend test suggested by Sasieni [1997]. Results from the Cochran-Armitage test can be combined across multiple studies via the generalized Cochran-Mantel-Haenszel test [Mantel, 1963]. If we assume that in J independent samples of cases and controls, individuals vary according to their doses of a treatment such that dose d ranges from 0, 1, 2, ..., D , then the counts of cases and controls receiving each dose within study j can be expressed as in Table I.

The generalized Cochran-Mantel-Haenszel statistic,

$$\frac{\left(\sum_{j=1}^J \left(\sum_{d=0}^D dr_{dj} - E\left(\sum_{d=0}^D dr_{dj}\right)\right)\right)^2}{\sum_{j=1}^J \text{Var}\left(\sum_{d=0}^D dr_{dj}\right)} = \frac{\left(\sum_{j=1}^J \left(\sum_{d=0}^D dr_{dj} - \frac{R_j}{N_j} \sum_{d=0}^D dn_{dj}\right)\right)^2}{\sum_{j=1}^J \frac{R_j S_j}{N_j - 1} \left(\frac{1}{N_j} \sum_{d=0}^D d^2 n_{dj} - \left(\frac{1}{N_j} \sum_{d=0}^D dn_{dj}\right)^2\right)},$$

where counts are defined as in Table I, has an asymptotic χ_1^2 distribution under the null hypothesis of no association. Note that this statistic is equivalent to the original Cochran-Armitage test statistic when $J = 1$, to the genotypic trend test when $J = 1$ and $D = 2$, and to a traditional Mantel-Haenszel test when $J > 1$ and $D = 1$. To combine the results of several genetic association studies, we are typically interested in the case where $J > 1$ and $D = 2$.

TABLE I. Counts of cases and controls in study j receiving possible doses of a treatment

	No. individuals receiving dose d					Total
	0	1	2	...	D	
# cases	r_{0j}	r_{1j}	r_{2j}		r_{Dj}	R_j
# controls	s_{0j}	s_{1j}	s_{2j}		s_{Dj}	S_j
Total	n_{0j}	n_{1j}	n_{2j}		n_{Dj}	N_j

Although the genotypic trend test is often used in case-control genetic association studies, it is not suitable for many studies. Studies that include quantitative outcome variables, continuous genotype scores (e.g. probabilistic scores for imputed SNP genotypes), additional covariates, or different statistical models cannot be summarized with simple genotype counts. In these situations, meta-analysis techniques that rely on basic summary statistics such as test statistics and their associated variance estimates may be useful. The inverse-variance approach [Woolf, 1955] uses a test statistic formed by taking the sum of point estimates (e.g. regression coefficients or odds ratios) divided by their estimated variance. Alternatively, normally distributed test statistics from different studies can be combined as a sample-size-weighted sum, where the weights are the square root of the ratio of each sample size to the total.

MULTIPLE TESTING ADJUSTMENT OF META-ANALYSES

For subject i , define \mathbf{Y}_i as a vector of K traits, \mathbf{G}_i as a vector of genotype scores for M SNP markers, and \mathbf{X}_i as a vector of other covariates. If we test the K traits for association with the M SNPs in a sample of N individuals via linear regression, logistic regression, or a Cochran-Armitage test for trend, the KM -length vector of score statistics is $\mathbf{U} = \sum_{i=1}^N (\mathbf{Y}_i - \bar{\mathbf{Y}}) \otimes \mathbf{G}_i$, where $\bar{\mathbf{Y}}$ is a vector of predicted trait values based on \mathbf{X}_i alone and \otimes represents the Kronecker product between two matrices. If we standardize the score statistics by dividing by the square root of their variance \mathbf{V} , the vector of standardized score statistics has an asymptotic multivariate normal distribution under the null hypothesis. For $L = KM$ tests, $((U_1/\sqrt{V_1}), \dots, (U_L/\sqrt{V_L})) \sim N(0, \mathbf{R})$, where \mathbf{R} can be estimated as the correlation matrix corresponding to the Kronecker product of the sample covariance of traits and the sample covariance of genotype score conditioned on covariates: $\text{Cov}(\mathbf{Y}) \otimes [\mathbf{G}\mathbf{G}^T - \mathbf{G}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{G}^T]$ [Conneely and Boehnke, 2007]. It is then possible to adjust for the L tests through integration of the multivariate normal distribution, as discussed below.

We can apply this result to groups of standard normal $(N(0,1))$ meta-analysis test statistics as well. Any meta-analysis test statistic that is a weighted sum of J standard normal statistics will itself follow a standard normal asymptotic distribution:

$$T = \frac{\sum_{j=1}^J w_j Z_j}{\sqrt{\sum_{j=1}^J w_j^2}} \sim N(0, 1). \tag{1}$$

As shown in Table II, the commonly used meta-analysis test statistics discussed above can be rewritten in this form.

TABLE II. Some common meta-analysis test statistics that can be written as in Equation (1)

Meta-analysis approach	Z_i	w_i
Cochran-Mantel-Haenszel	$\frac{\sum_{d=0}^D dr_{dj} - E(\sum_{d=0}^D dr_{dj})}{\sqrt{\text{Var}(\sum_{d=0}^D dr_{dj})}}$	$\sqrt{\text{Var}(\sum_{d=0}^D dr_{dj})}$
Inverse variance method (based on point estimate θ_j)	$\frac{\theta_j - E(\theta_j)}{\sqrt{\text{Var}(\theta_j)}}$	$\frac{1}{\sqrt{\text{Var}(\theta_j)}}$
Weighted sum method	Any $N(0, 1)$ test statistic	$\sqrt{\frac{N_j}{\sum_{i=1}^J N_i}}$

A vector of L meta-analysis test statistics taking the form in (1) is a vector-weighted sum of J multivariate normal vectors of single-study test statistics:

$$\mathbf{T} = \left(\frac{\sum_{j=1}^J w_{1j} Z_{1j}}{\sqrt{\sum_{j=1}^J w_{1j}^2}}, \dots, \frac{\sum_{j=1}^J w_{Lj} Z_{Lj}}{\sqrt{\sum_{j=1}^J w_{Lj}^2}} \right).$$

Under the null hypothesis of no association, \mathbf{T} will be distributed multivariate normally with mean vector zero and

$$\text{Cov}(T_k, T_l) = \frac{\sum_j w_{kj} w_{lj} R_{kl}^{(j)}}{\sqrt{\sum_{j=1}^J w_{kj}^2 \sum_{j=1}^J w_{lj}^2}}. \tag{2}$$

where $R_{kl}^{(j)}$ represents element k, l of the correlation matrix between tests for study j , $\mathbf{R}^{(j)}$. If L is small relative to sample size, the empirical correlation matrix for study j will provide a good approximation to $\mathbf{R}^{(j)}$, but the quality of this approximation is decreased as L increases relative to sample size. Due to the high dimensionality of the covariance matrices to be estimated and variation in sample size between studies, we will estimate all correlation matrices using the shrinkage estimators of Schäfer and Strimmer [2005] as implemented in the R package `corpcor`.

For a given realization of $\mathbf{T} = [T_1, T_2, \dots, T_L]$, we can compute a P -value for the most extreme test statistic $T_{\max} = \max_{1 \leq j \leq L} |T_j|$ as $P_{\text{ACT}} = 1 - P(-T_{\max} < W_1, W_2, \dots, W_L < T_{\max})$, where W_1, W_2, \dots, W_L are random variates from a multivariate normal distribution with covariance defined as in equation (2) [Conneely and Boehnke, 2007]. P_{ACT} (P -value adjusted for correlated tests) can be computed as a multivariate normal integral. As we have previously [Conneely and Boehnke, 2007], we will use the computationally efficient method of numerical integration developed by Genz [1992] as implemented in the `mvtnorm` package [Genz et al., 2007] for R [R Development Core Team, 2008]. The `mvtnorm` package can perform numerical integration for $L \leq 1,000$, although we have previously recommended computing P_{ACT} for 500 or fewer tests to avoid imprecise estimates of the correlation matrix \mathbf{R} [Conneely and Boehnke, 2007]. This will not be a major limitation in the context of most large-scale candidate gene association studies. For roughly independent genes, a gene-specific P_{ACT} can be computed separately for each gene based on the correlation matrix for SNPs within that gene and the most extreme test statistic experiment-wide (T_{\max}). P_{ACT} can then be adjusted for the number of genes tested via a Bonferroni-like correction that is analogous to the method of Šidák [1967]: $P_{\text{ACT}} = 1 - \prod_{g=1}^G (1 - P_g)$,

where G is the number of genes tested and P_g is the gene-specific estimate of P_{ACT} for gene g .

TREATMENT OF MISSING DATA

Estimation of the covariance matrix in (2) is straightforward when all genotypes and traits are observed for all individuals. To estimate P_{ACT} in the presence of missing data at the individual level, we have previously suggested that each association test be performed using all available observations, but setting missing genotypes or phenotypes to their mean values for purposes of covariance estimation [Conneely and Boehnke, 2007]; this ensures that the covariance matrix is positive definite and is equivalent to setting to zero the individual component of the score statistic for missing observations. In the context of genetic association studies, this choice requires estimating the variance on the full data set where missing values for trait k have been set to the mean value for trait k and missing values for marker m have been set to the mean genotype score for marker m .

A similar approach can be applied in a meta-analysis framework when the complete set of traits or markers is not necessarily available in every study. If the availability of traits or markers in certain studies is independent of results in other studies (e.g. if genotypes for a particular marker are missing in certain studies due to assay failure or limited funding), unavailable traits or markers can simply be treated as missing data. If test l is not performed in study j , then the meta test statistics and variance can be computed with Z_{ij} and w_{ij} set to zero for all individuals in sample j , since this test is missing-at-random for these individuals. This allows computation of the meta test statistics and covariance estimates based only on studies with data for the relevant traits and markers while ensuring a positive-definite covariance matrix.

TWO-STAGE DESIGN

A different approach is required if availability of markers and/or traits in one study is dependent on test statistics from another study, since in this situation the missingness of genotype and/or trait data is not random. The two-stage design, another type of meta-analysis commonly employed in genetic association studies, involves the genotyping of many markers in an initial study, followed by the genotyping of only a subset of markers whose test statistics exceed a pre-determined cutoff in an additional study. For Z -statistics with $N(0,1)$ distributions under the null hypothesis of no association, Skol et al. [2006, 2007] showed that if test statistic z_1 is observed in an initial study and only SNPs for which $|z_1| > T_1$ are tested in a replication study, the conditional probability of the weighted meta-statistic $z_{\text{joint}} = \sqrt{(N_1/N)}z_1 + \sqrt{(N_2/N)}z_2$ reaching significance, $P_{\text{joint}} = P(|z_{\text{joint}}| > T_{\text{joint}} | |z_1| > T_1)$, can be obtained through integration of the conditional normal cumulative distribution function. The overall P -value for the joint analysis is then $P_1 P_{\text{joint}}$, where $P_1 = P(|z_1| > T_1)$ and P_1 and P_{joint} are both computed under the null hypothesis of no association.

For a two-stage meta-analysis involving L correlated tests, P_{ACT} can be used conditionally to adjust for the correlation between tests while taking the conditional selection of tests into account. The appropriate adjusted P -value for the best observed meta-statistic is the joint probability that under the null hypothesis, at least one of the L tests would (1) pass the predetermined cutoff T_1 in

the initial study and (2) equal or exceed the best observed meta-statistic T_{joint} :

$$P_{ACT-2s} = P(|z_{1,1}| > T_1, |z_{l,\text{joint}}| > T_{\text{joint}} \text{ for some } l \in \{1, 2, \dots, L\}) \quad (3)$$

Defining the L initial test statistics as $z_{1,1}, z_{2,1}, \dots, z_{L,1}$ and the L joint test statistics (of which some or all will be unobserved) as $z_{1,\text{joint}}, z_{2,\text{joint}}, \dots, z_{L,\text{joint}}$, the set of $2L$ initial and joint statistics has a multivariate normal distribution with covariance matrix

$$\begin{bmatrix} \mathbf{R}^{(1)} & \frac{w_1 \mathbf{R}^{(1)}}{\sqrt{\sum w_j^2}} \\ \frac{w_1 \mathbf{R}^{(1)}}{\sqrt{\sum w_j^2}} & \frac{\sum w_j^2 \mathbf{R}^{(j)}}{\sum w_j^2} \end{bmatrix}$$

where $\mathbf{R}^{(j)}$ and w_j are the sample correlation matrices and weights for study j as defined above. P_{ACT-2s} (Equation (3)) can then be computed as a piecewise sum of multivariate normal probabilities or a much less computationally intensive approximation that yields near-identical results (see Appendix A for details).

SIMULATIONS

To assess the validity and power of our method of adjusting meta-statistics for multiple testing, we simulated candidate gene association tests in an initial study and five replication studies. In each simulation, we randomly drew individual genotypes for 59 SNPs covering a 1-Mb region that included the lactase gene (*LCT*). To simulate variation between studies, we drew the individual genotypes from six different collections of subjects from The Population Reference Sample (POPRES) [Nelson et al., 2008]; this allowed the correlation between SNPs, and hence between SNP association tests, to vary between studies as it would in a meta-analysis of studies from highly diverse populations. The POPRES data were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000145.v2.p2 through dbGaP accession number phs000145.v2.p2. Table III shows some characteristics of the six simulated studies and the respective POPRES subject collections from which they were drawn.

For each simulated individual, we simulated four correlated traits (two binary and two quantitative) with frequencies and means varying between studies. To obtain

TABLE III. POPRES study collections used in simulations

POPRES study collection	No. genotyped QC+ subjects available	No. sampled in each simulation	HapMap counterpart
CoLaus (Switzerland)	2,507	300	CEU
UCSF African American	346	300	ASW
Japanese	73	60	JPT
Mexican	112	100	MEX
LOLIPOP: Indian Asians	359	325	GIH
LOLIPOP: European Caucasians	481	325	CEU

correlated traits, we first simulated four continuous traits such that for individual i in study j , trait $k = Y_{ij}^{(k)} = Z_{ij}^{(k)} + \mu_j + \beta^{(k)}(G_{ij}^{(m)} - \bar{G}_{ij}^{(m)})$, where $Z^{(1)}, \dots, Z^{(4)}$ are random variates from the multivariate normal distribution, $\beta^{(k)}$ is a constant effect size that is 0 except in power calculations, and $G_{ij}^{(m)}$ is the number of copies of the risk allele of SNP m possessed by individual i , with mean $\bar{G}_{ij}^{(m)}$ in study j . We then defined the quantitative traits as $Y^{(1)}$ and $Y^{(2)}$, and the binary traits as 1 if $Y^{(k)} > 0$ and 0 otherwise for $k = 3, 4$. Traits were simulated so that the correlation between each pair of traits was approximately 0.7.

To assess type I error for each of the above study designs, we performed 5,000 simulations, where all four traits were simulated independently of genotype (i.e., $\beta^{(k)} = 0$ for $k = 1, \dots, 4$), and we compared the 5,000 multiple-testing-adjusted P -values to the expected sample quantiles. To assess power, we created 1,000 simulation replicates where one of the binary traits was influenced by a single SNP. We estimated power as the proportion of 1,000 simulations in which the adjusted P -value < 0.05 . We performed simulations to assess the rates of type I error and power for three types of meta-analysis, representing a prospective, retrospective, and two-stage meta-analysis.

Prospective meta-analysis. In each of the six studies, we tested each of the 59 SNPs for association with each of the four traits, for a total of 236 unique tests. Depending on whether each trait was binary or quantitative, we used either a Cochran-Armitage test or linear regression to test for association between the trait and the number of copies the minor allele for each SNP (0, 1, or 2). We combined each of the 236 test statistics across six studies with one of two meta-analysis approaches. First, we performed a sample-size weighted meta-analysis where the Z statistic for each study was either the signed square root of the χ^2 statistic from the Cochran-Armitage test (binary traits), or the t -statistic from a linear regression of the trait value on the number of allele copies (quantitative traits). Second, we used a Cochran-Mantel-Haenszel test to combine tests that involved the binary traits, and an inverse-variance meta-analysis based on the linear regression coefficient to combine tests that involved the continuous traits across the six studies. To adjust for multiple correlated tests, we computed the 236×236 correlation matrix for the meta-analysis as the weighted sum of the within-study correlation matrices as in Equation (2), using the weights shown in Table II. To simulate an extreme case of the scenario where the set of SNPs does not fully overlap across studies, a large number of SNPs (~20%) were missing in each study for random reasons; the weights and correlations involving these tests were set to zero as described above. We then computed P_{ACT} to obtain a multiple-testing-adjusted P value.

Retrospective meta-analysis. We performed the same meta-analysis as in the preceding paragraph, but here we assumed that (1) only summary statistics were available for each test in each study, (2) the correlation matrix between the traits was available for the initial study only, and (3) correlation matrices between the SNPs could not be computed for any study. To estimate the SNP correlation matrices for each study, we used genotypes from the closest available Phase 3 HapMap sample for each POPRES study collection, as shown in the rightmost column of Table III. For each study, we computed the

overall correlation matrix for the 236 tests as the Kronecker product of the HapMap correlation matrices and the trait correlation matrix from the initial study. We computed the weighted correlation matrix based on equation (2), and computed P_{ACT} based on this correlation matrix.

Two-stage meta-analysis. We assumed a two-stage design where all 236 SNP-trait combinations were tested in an initial study (CoLaus, described in first row of Table III), but only SNP-trait combinations with individual association test P -values of < 0.1 were tested in the five follow-up studies. In practice, the number of tests passing the criteria for further testing in each simulation under the null hypothesis ranged from 0 (in 0.5% of simulations) to 131, with a median of 19 tests followed up. For those tests selected for follow-up in all studies, we computed meta test statistics and adjusted the most extreme meta test statistic for multiple correlated tests and conditional selection for follow-up with P_{ACT} as in equation (3), using the approximation described in Appendix A.

RESULTS

We performed simulated meta-analyses for a variety of study designs in which four traits were tested for association with 59 SNPs in the vicinity of the *LCT* gene across six studies.

Adjusted P -values for several simulated meta-analyses are plotted on a \log_{10} scale against their theoretical quantiles in Figure 1A–C. Figure 1A and B reflect the prospective and retrospective study designs described in Methods in which all four traits were tested for association with all available SNPs in each study, and results were combined via either the Cochran-Mantel-Haenszel test (for binary traits), or the inverse-variance method (for quantitative traits). For each simulation, we adjusted the best meta P -value in each simulation for multiple testing with P_{ACT} . We carried out the adjustment assuming either that individual-level data were available and could be used to compute study-specific correlation matrices (Fig. 1A), or that only summary-level data were available (Fig. 1B), in which case we used HapMap correlation matrices as proxies to estimate study-specific correlation matrices. In general, values of P_{ACT} fall within the 95% confidence bounds and follow the identity line quite closely for the entire range of P -values, indicating that the appropriate type-I error rate is maintained at all levels of significance.

Figure 1C reflects a two-stage meta-analysis, where all tests with association P -values of < 0.10 in the initial study were tested in the five follow-up studies. As in Figure 1A and B, results were combined across studies via either the Cochran-Mantel-Haenszel test or the inverse-variance method. We computed P_{ACT} for each simulation using the approximation to equation (3) described in Appendix A and plotted the values against their quantiles on a \log_{10} scale. P_{ACT} once again tracks its quantiles quite closely, indicating that it achieves the correct type-I error rate for all reasonable α -levels.

A comparison of the power of the methods described above is presented in Table IV for 1,000 simulated meta-analyses based on Cochran-Mantel-Haenszel and inverse-variance test statistics. By accounting for the correlation between tests, adjustment for multiple testing with P_{ACT} leads to gains in power over Bonferroni adjustment. Notably, little power is lost when study-specific correlation

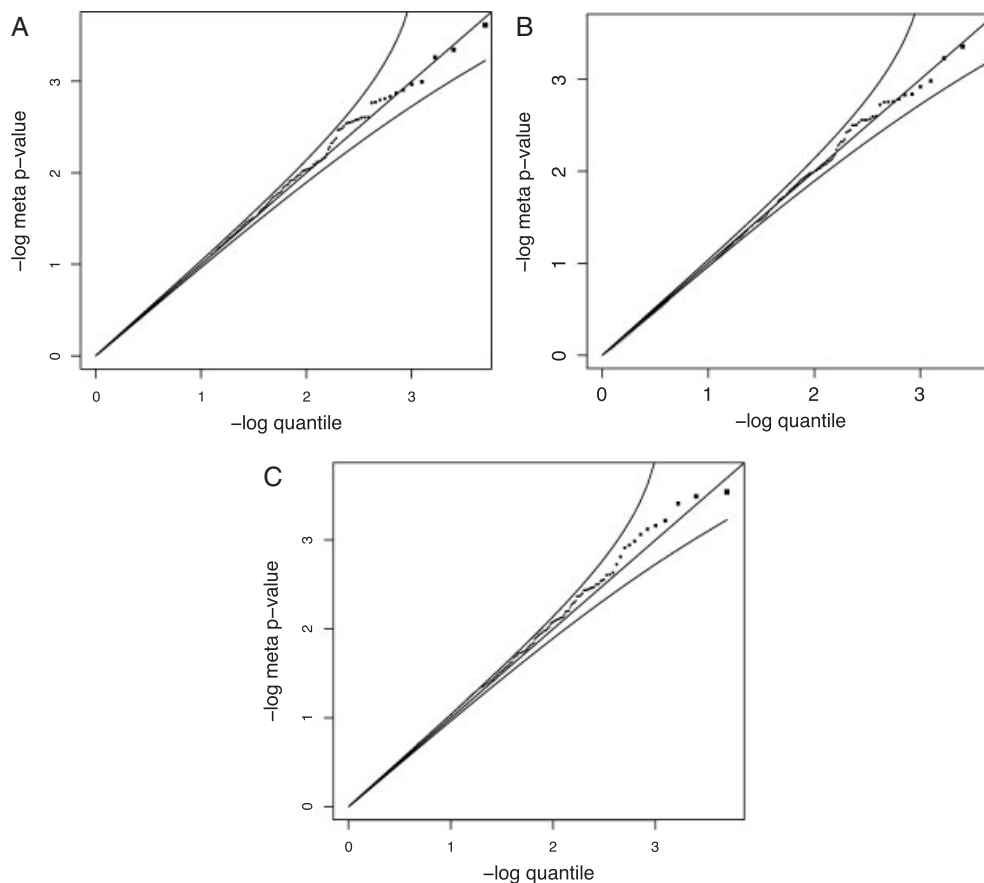


Fig. 1. $-\log_{10}$ of P_{ACT} -adjusted P -values plotted against theoretical quantiles (diagonal line) with 95% confidence intervals (curved lines). P -values are from 5,000 simulated meta-analyses where 59 SNPs were tested for association with four traits (A) in six studies under a prospective design where individual-level data were available, or (B) in six studies, under a retrospective design where only summary statistics were available, or (C) in an initial study, under a two-stage design where only tests with $P < 0.10$ were followed up in five additional studies.

TABLE IV. Estimated power to detect a heterogeneous genetic association between a SNP and a binary trait, based on 1,000 simulated meta-analyses of 59 SNPs and four traits

Type of study performed	Multiple testing adjustment		
	Bonferroni	P_{ACT}	P_{ACT} based on HapMap
Full meta-analysis of six studies			
All 236 tests performed in all studies	0.72	0.79	0.78
~20% of SNPs missing in each study	0.68	0.73	0.72
Two-stage analysis	0.58	0.69	0.68

matrices are estimated from the HapMap populations rather than the individual-level study data.

Although the results in Figure 1A–C and Table IV are all based on the Cochran-Mantel-Haenszel and inverse-variance meta-analysis approaches, we obtained very similar results using a sample-size-weighted approach (results not shown).

To illustrate our method, we applied it to a recently published meta-analysis. In a two-stage analysis involving

116 candidate genes, Wang et al. [2010] tested 894 tag SNPs for association with coronary atherosclerosis in the Han Chinese population. After follow-up of 51 SNPs with $P < 0.05$ in their stage I study ($N = 586$), they found that three SNPs in *ITGA2*, *PON1*, and *THBS2* had P -values of < 0.05 in both the stage I study and the stage II study ($N = 1,794$). Since the meta-analysis P -values for the three SNPs ($P = 9.2 \times 10^{-5}$, 1.9×10^{-4} , and 3.0×10^{-3}) exceeded their chosen Bonferroni significance threshold of $0.05/894 = 5.6 \times 10^{-5}$ for the 894 SNPs, the evidence for association could be classified as merely suggestive rather than significant. However, the Bonferroni approach does not take into account the correlation between SNPs within genes or the two-stage design. To address this issue, we used genotype data from the CHB (Han Chinese in Beijing) HapMap sample to estimate correlation matrices for SNPs in each of the 116 genes using the shrinkage estimators of Schäfer and Strimmer [2005]. Since CHB genotype data were available for 726 of the 894 SNPs, we treated the remaining 168 SNPs as though they were uncorrelated with all others, which should have made our estimates of $P_{\text{ACT}-2s}$ slightly conservative. We computed $P_{\text{ACT}-2s}$ for the generalized Cochran-Mantel-Haenszel test as described in Methods, by first computing an estimate of $P_{\text{ACT}-2s}$ for each gene, and then adjusting for the number of genes via the modified version of the Sidák [1967] procedure shown

in Methods. While adjustment for multiple tests via the standard Šidák [1967] procedure yielded adjusted P -values of 0.079, 0.16, and 0.93 for the top three SNPs, our estimate of $P_{\text{ACT-2s}}$ was 0.042 for the top SNP ($P_{\text{ACT-2s}} > 0.05$ for the other two SNPs), so that this SNP, from *ITGA2*, attained experiment-wide significance in the two-stage analysis.

DISCUSSION

We have presented a new approach for adjustment of multiple correlated association tests in meta-analyses that can be applied to any number of independent studies, and have made software available at http://csg.sph.umich.edu/boehnke/p_act.php. In simulations of 236 correlated association tests analyzed across 6 studies, our methods attained the appropriate type I error rates for a range of study designs and provided a gain in power over Bonferroni-style adjustments, which do not account for correlation between tests. Our approach remains valid and powerful even when all SNPs are not genotyped in all studies, when only select SNPs are followed up in additional studies, and when individual-level data are not available for analysis.

A potential limitation of our analysis is that we restricted our focus to fixed effects, rather than random effects, approaches to meta-analysis. This decision was guided by the fact that even for the largest consortia, the number of studies combined is generally too small for the between-study variance to be accurately estimated or for the assumption of asymptotically normal sample effects to be realistic. Pfeiffer et al. [2009] recently evaluated the performance of both fixed and random effects meta-analysis approaches in a variety of settings, and observed that the fixed effects approach performed as well as or better than the random effects approach in all settings considered.

Since the methods we suggest here are based on estimation of large covariance matrices and numerical integration of the multivariate normal distribution, there are practical limitations on the number of tests that can be included in a single analysis. The R package *mvtnorm* [Genz et al., 2007] is capable of computing multivariate normal integrals of dimension up to 1000; however, we have previously observed that P_{ACT} works best for 500 or fewer tests [Conneely and Boehnke, 2007]. One issue is that the precision with which we can estimate the covariance matrix suffers as the number of parameters overtakes the sample size. To address the precision issues inherent in estimating covariance matrices of large dimension based on moderate sample sizes (or small sample sizes when HapMap samples are used as proxies), we employ the shrinkage estimators of Schäfer and Strimmer [2005].

Because the number of tests is limited to hundreds, rather than the hundreds of thousands of tests performed routinely in GWAS, the methods presented here are best suited for collaborative candidate gene studies rather than genome-wide approaches. However, candidate gene studies may be in a position to benefit the most from new methods of multiple-testing adjustment. As the density of SNP coverage in GWAS studies continues to increase due to the availability of larger arrays and methods for imputing SNPs not present on the arrays, the common approach of adopting a genome-wide significance threshold that reflects all of the potential testable hypotheses may be the best solution (see, for example, Dudbridge and Gusnanto [2008]). On the other hand, candidate gene studies remain targeted efforts with discrete hypotheses that

require study-specific significance thresholds. The methods we present here can be applied to the full spectrum of collaborative candidate gene studies, ranging from studies of a single gene to large-scale studies testing thousands of genes, and can be applied in situations where permutation testing may not be feasible, due to lack of individual data or lack of data on SNPs not selected for follow-up in two-stage studies.

Due to a wealth of available data and the drive to combine information as a means of affirming valid results and ruling out spurious ones, meta-analyses have become increasingly common in the genetic association literature. Given this current emphasis and what will likely be a continued focus, we feel that the development of methods to integrate meta-analysis techniques into genetic association studies are timely and have the potential to be useful in a variety of settings.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Health (NIH) grant HG000376 (to M.B.). This work benefitted from the helpful comments of an anonymous reviewer.

REFERENCES

- Armitage P. 1955. Tests for linear trends in proportions and frequencies. *Biometrics* 11:375–386.
- Cochran WG. 1954. Some methods for strengthening the common χ^2 tests. *Biometrics* 10:417–451.
- Conneely KN, Boehnke M. 2007. So many correlated tests, so little time! Rapid adjustment of P -values for multiple correlated tests. *Am J Hum Genet* 81:1158–1168.
- Dudbridge F. 2006. A note on permutation tests in multistage association scans. *Am J Hum Genet* 78:1094–1095.
- Dudbridge F, Gusnanto A. 2008. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 32:227–334.
- Genz A. 1992. Numerical computation of multivariate normal probabilities. *J Comput Graph Stat* 1:141–149.
- Genz A, Bretz F, Hothorn T. 2007. *mvtnorm: multivariate normal and t distribution*. R package version 0.8–0.
- Li Y, Willer C, Sanna S, Abecasis G. 2009. Genotype imputation. *Annu Rev Genomics Hum Genet* 10:387–406.
- Lin DY. 2006. Evaluating statistical significance in two-stage genome-wide association studies. *Am J Hum Genet* 78:505–509.
- Lin DY, Zeng D. 2009. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol* 32:60–66.
- Mantel N. 1963. Chi-square tests with 1 degree of freedom – extensions of the Mantel-Haenszel procedure. *J Am Stat Assoc* 58:690–700.
- Mantel N, Haenszel W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22:719–748.
- Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J, Briley LP, Maruyama Y, Waterworth DM, Waeber G, Vollenweider P, Oksenberg JR, Hauser SL, Stirnadel HA, Kooper JS, Chambers JC, Jones B, Mooser V, Bustamante CD, Roses AD, Burns DK, Ehm MG, Lai EH. 2008. The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* 83:347–358.
- Pearson K. 1904. Report on certain enteric fever inoculation statistics. *BMJ* 3:1243–1246.
- Pfeiffer RM, Gail MH, Pee D. 2009. On combining data from genome-wide association studies to discover disease-associated SNPs. *Stat Sci* 24:547–560.
- R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (<http://www.R-project.org>).

Sasieni PD. 1997. From genotypes to genes: doubling the sample size. *Biometrics* 53:1253–1261.

Schäfer J, Strimmer K. 2005. A shrinkage approach to large-scale covariance-matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* 4:Article 32.

Šidák Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc* 62:626–633.

Skol AD, Scott LJ, Abecasis GR, Boehnke M. 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38:209–213.

Skol AD, Scott LJ, Abecasis GR, Boehnke M. 2007. Optimal designs for two-stage genome-wide association studies. *Genet Epidemiol* 31:776–788.

Wang Y, Fu W, Xie F, Wang Y, Chu X, Wang H, Shen M, Wang Y, Wang Y, Sun W, Lei R, Yang L, Wu H, Foo J, Liu J, Jin L, Huang W. 2010. Common polymorphisms in *ITGA2*, *PON1* and *THBS2* are associated with coronary atherosclerosis in a candidate gene association study of the Chinese Han population. *J Hum Genet* 55:490–494.

Woolf B. 1955. On estimating the relation between blood group and disease. *Ann Human Genet* 19:251–253.

APPENDIX A

APPROXIMATION FOR P_{ACT-2s}

The probability $P_{ACT-2s} = P(|z_{l,1}| > T_1 \text{ and } |z_{l,joint}| > T_{joint})$ for at least one $l \in \{1, 2, \dots, L\}$ from equation (3) may be computed as 1 minus the probability that there is no $l \in \{1, 2, \dots, L\}$ for which both $|z_{l,1}| > T_1$ and $|z_{l,joint}| > T_{joint}$. This probability can be computed as a piecewise sum of 3^L probabilities. For example, when $L = 2$,

$$\begin{aligned}
 P_{ACT-2s} &= 1 - P(z_{1,1} > T_1, |z_{1,joint}| \leq T_{joint}, z_{2,1} > T_1, |z_{2,joint}| \leq T_{joint}) \\
 &\quad - P(z_{1,1} < -T_1, |z_{1,joint}| \leq T_{joint}, z_{2,1} > T_1, |z_{2,joint}| \leq T_{joint}) \\
 &\quad - P(|z_{1,1}| \leq T_1, z_{2,1} > T_1, |z_{2,joint}| \leq T_{joint}) \\
 &\quad - P(z_{1,1} > T_1, |z_{1,joint}| \leq T_{joint}, z_{2,1} < -T_1, |z_{2,joint}| \leq T_{joint}) \\
 &\quad - P(z_{1,1} < -T_1, |z_{1,joint}| \leq T_{joint}, z_{2,1} < -T_1, |z_{2,joint}| \leq T_{joint}) \\
 &\quad - P(|z_{1,1}| \leq T_1, z_{2,1} < -T_1, |z_{2,joint}| \leq T_{joint}) \\
 &\quad - P(z_{1,1} > T_1, |z_{1,joint}| \leq T_{joint}, |z_{2,1}| \leq T_1) \\
 &\quad - P(z_{1,1} < -T_1, |z_{1,joint}| \leq T_{joint}, |z_{2,1}| \leq T_1) \\
 &\quad - P(|z_{1,1}| \leq T_1, |z_{2,1}| \leq T_1)
 \end{aligned}$$

Given that

$$\begin{aligned}
 &[z_{1,1}, z_{2,1}, \dots, z_{L,1}, z_{1,joint}, z_{2,joint}, \dots, z_{L,joint}]^T \\
 &\sim N \left(\mathbf{0}, \begin{bmatrix} \mathbf{R}^{(1)} & \frac{w_1 \mathbf{R}^{(1)}}{\sqrt{\sum w_j^2}} \\ \frac{w_1 \mathbf{R}^{(1)}}{\sqrt{\sum w_j^2}} & \frac{\sum w_j^2 \mathbf{R}^{(j)}}{\sum w_j^2} \end{bmatrix} \right),
 \end{aligned}$$

each of these probabilities can be computed numerically based on the multivariate normal distribution. However, the computation of 3^L separate probabilities is feasible only for small L . Alternatively, if the between-test correlation matrices can be assumed to be similar across studies such that $\frac{\sum w_j^2 \mathbf{R}^{(j)}}{\sum w_j^2}$ can be approximated with $\mathbf{R}^{(1)}$, a good approximation to (3) is available.

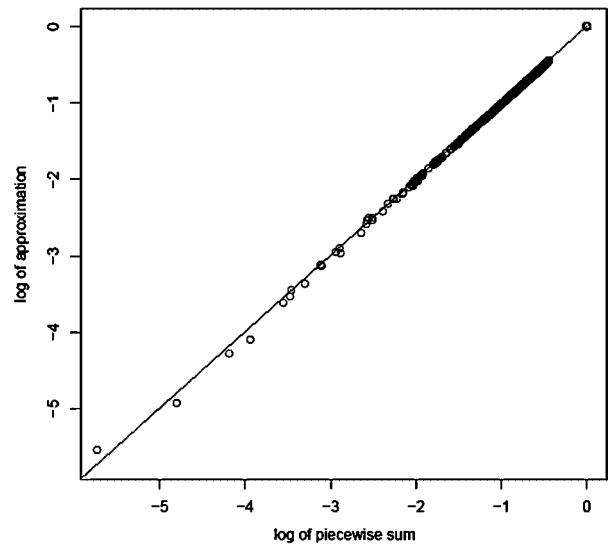


Fig. A1. Comparison of P_{ACT-2s} estimated for 8 correlated tests as either a piecewise sum of probabilities or a faster 2-step approximation.

To perform this approximation, we adjust the minimum P -value P_{min} from the two-stage meta-analysis in two steps. We first adjust for the two-stage test by computing the probability $P' = P(|z_1| > T_1, |z_{joint}| > |T_{joint}|)$ that a single test passes both the initial cutoff in the first study and attains the magnitude of the best test statistic observed in the combined studies, T_{joint} . Using the fact that the joint distribution of z_1 and z_{joint} is bivariate normal with correlation w_1 , P' can be easily computed as the sum of four probabilities:

$$\begin{aligned}
 &P(z_1 > T_1, z_{joint} > |T_{joint}|) + P(z_1 > T_1, z_{joint} < -|T_{joint}|) \\
 &\quad + P(z_1 < -T_1, z_{joint} > |T_{joint}|) \\
 &\quad + P(z_1 < -T_1, z_{joint} < -|T_{joint}|)
 \end{aligned}$$

By computing the probability P' , we have computed a P -value for the best test statistic that reflects the joint probability that (1) for a single test, the magnitude of the stage 1 test statistic exceeds the cutoff T_1 and (2) the meta-analysis statistic attains the magnitude of T_{joint} . Thus, P' has been adjusted for the additional burden of the stage 1 test statistic having to exceed T_1 , but has not yet been adjusted for the multiple tests that were performed in the first stage. We can adjust for the L tests in a second step by transforming P' to a Z -score $Z' = \Phi^{-1}(1 - P'/2)$ and adjusting Z' for the L tests that were performed by computing P_{ACT} assuming L tests with correlation matrix estimated as in equation (2), where weight w_{ij} is set to 0 if test l was not performed in study j .

To test the performance of the approximation, we computed P_{ACT-2s} using both the exact method and the approximation described above for 1,000 simulations similar to those presented in Results involving $L = 8$ correlated tests. As Figure A1 shows, we obtained near-identical results with the two methods, demonstrating that the approximation is highly accurate in a situation with heterogeneous samples of individuals and high correlation between tests.