Meta-Analysis of Gene Level Tests for Rare Variant Association

Published in final edited form as:
Nat Genet. 2014 February ; 46(2): 200–204. doi:10.1038/ng.2852.

Meta-Analysis of Gene Level Tests for Rare Variant Association

Dajiang J. Liu[1,*], **Gina M. Peloso**[2,3,4,*], **Xiaowei Zhan**[1,*], **Oddgeir L. Holmen**[5,6,*], **Matthew Zawistowski**[1], **Shuang Feng**[1], **Majid Nikpay**[7], **Paul L. Auer**[8,9], **Anuj Goel**[10,11], **He Zhang**[12,13], **Ulrike Peters**[8,14], **Martin Farrall**[10,11], **Marju Orho-Melander**[11,15], **Charles Kooperberg**[8,16], **Ruth McPherson**[7], **Hugh Watkins**[10,11], **Cristen J. Willer**[12,13], **Kristian Hveem**[5,17], **Olle Melander**[11,15], **Sekar Kathiresan**[2,3,4,18,+], and **Gonçalo R. Abecasis**[1,+]

[1]Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI 48109

[2]Broad Institute of Harvard and MIT, Cambridge, MA

[3]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA

[4]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA

[5]Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim 7489, Norway

[6]St. Olav Hospital, Trondheim University Hospital, Trondheim, Norway

[7]University of Ottawa Heart Institute, Ottawa, Ontario, Canada

[8]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle WA 98109, USA

[9]School of Public Health, University of Wisconsin-Milwaukee

[10]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom

[11]Department of Cardiovascular Medicine, University of Oxford, Oxford, UK

[12]Division of Cardiology, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109

[13]Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109

[14]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA

[15]Department of Clinical Sciences, Lund University, Malmö, Sweden

[16]Department of Biostatistics, University of Washington School of Public Health, Seattle, WA

Correspondence To: Dajiang J. Liu (dajiang@umich.edu) Or Gonçalo R. Abecasis (goncalo@umich.edu) Center for Statistical Genetics, Department of Biostatistics, 1415 Washington Heights, Ann Arbor, MI 48109.
*These authors contributed equally and should be considered joint first authors.
+These authors jointly directed the study.

URLs: Author's website at http://genome.sph.umich.edu/wiki/RAREMETAL-SOFTWARE

Author Contributions: D.J.L., S.K. and G.R.A. conceived and designed the study. D.J.L., G.M.P., X.Z. carried out primary data analysis. D.J.L., X.Z. and S.F. wrote the software package implementing proposed methodologies. O.L.H., M.N., P.L.A., A.G., H.Z., U.P., M.F., M.O., C.K., R.M., H.W., C.J.W. K.H., O.M. contributed phenotypes, exome array genotypes, and analyses for the study. M.Z. conducted population genetics simulation analysis. D.J.L. and G.R.A. wrote the first version of the manuscript. All authors critically reviewed and approved the manuscript. S.K. and G.R.A. jointly supervised the study.

[17]Levanger Hospital, Levanger, Norway

[18]Harvard Medical School, Cambridge, MA

Abstract

The vast majority of connections between complex disease and common genetic variants were identified through meta-analysis, a powerful approach that enables large sample sizes while protecting against common artifacts due to population structure, repeated small sample analyses, and/or limitations with sharing individual level data. As the focus of genetic association studies shifts to rare variants, genes and other functional units are becoming the unit of analysis. Here, we propose and evaluate new approaches for performing meta-analysis of rare variant association tests, including burden tests, weighted burden tests, variable threshold tests and tests that allow variants with opposite effects to be grouped together. We show that our approach retains useful features of single variant meta-analytic approaches and demonstrate its utility in a study of blood lipid levels in ∼18,500 individuals genotyped with exome arrays.

Introduction

Proceeding from the discovery of a genetic association signal to a mechanistic insight about human biology should be much easier for one or a set of alleles with clear functional consequence, including non-synonymous, splice altering and protein truncating alleles. Most of these alleles are very rare, with only one such allele expected to reach MAF>5% in the average human gene[1]. Recent advances in exome sequencing and the development of exome genotyping arrays are enabling explorations of the very large reservoir of rare coding variants in humans and are expected to accelerate the pace of discovery in human genetics[2].

Rare variants can be examined using association tests that group alleles in a gene or other functional unit[3]. Compared to tests of individual alleles, this grouping can increase power, especially when applied to large samples where several rare variants are observed in the same functional unit[4]. The simplest rare variant tests consider the number of potentially functional alleles in each individual[5], but the tests can be refined to weigh variants according to their likely functional impact[6], to allow for imputed or uncertain genotypes[7,8], or to allow variants that increase and decrease risk to reside in the same gene[9-11] (a feature that is important when the same gene harbors hypermorph and hypomorph alleles[12]). The optimal strategy for grouping and weighting rare variants – ranging from focusing on protein truncation alleles to examining all non-synonymous variants and encompassing strategies that examine all variants with frequency <5% as well as alternatives that examine only singletons – depends on the unknown genetic architecture of each trait and each locus[13].

Here, we describe practical approaches for meta-analysis of rare variants. Our approach starts with simple statistics that can be calculated in an individual study (single site score statistics and their covariance matrix, which summarizes the linkage disequilibrium information and relatedness among sampled individuals). We then show that, when these statistics are shared, a wide variety of gene-level association tests can be executed centrally – including both weighted or un-weighted burden tests with fixed[5] or variable frequency threshold[6] and sequence kernel association tests (SKAT) that accommodate alleles with opposite effects within a gene[9]. Our approach generates comparable results to sharing individual level data (and, in fact, identical results when allowing for between study heterogeneity in nuisance parameters, such as trait means, variances and covariate effects). As an illustration of our approach, we analyze blood lipid levels in >18,500 individuals genotyped with exome genotyping arrays. Our analysis of blood lipid levels provides examples of loci where signal for gene-level association tests exceeds signal for single variant tests and shows that our approach can recover signals driven by very rare variants

(frequency <0.05%). Given that very large sample sizes are required for successful rare variant association studies, we expect our methods (and refined versions thereof) will be widely useful.

Our approach is based on the insight that analogues of most gene level association tests can be constructed using single variant test statistics and knowledge of their correlation structures. As shown in **Methods**, simple[14] and weighted[10,15] burden tests, variable threshold tests[6] and tests allowing for variants with opposite effects[9] can be constructed in this manner. We meta-analyze single variant statistics using the Cochran-Mantel-Haenszel method, calculate variance-covariance matrices for these statistics, and construct gene-level association tests by combining the two. In Supplementary Notes, we show that rare variant statistics generated in this way are identical to those obtained by sharing individual level data and allowing for heterogeneity in nuisance parameters, with no loss of power. Importantly, rare variant statistics calculated in this way are less vulnerable to artifacts due to population stratification than statistics generated by naïvely pooling individual level data. As in other meta-analysis settings, sharing summary statistics accelerates the overall analysis process, mitigates concerns about participant confidentiality, and reduces the risk that data will be used for unapproved analyses (as always, to avoid violating the trust of research subjects, we strongly recommend that investigators sharing summary statistics agree that these will not be used to identify research subjects). For evaluating significance, we propose methods for calculating p-values using asymptotics and also Monte-Carlo methods that use knowledge of linkage disequilibrium relationships to sample plausible combinations of single variant statistics and then generate empirical distributions for our gene-level statistics. Since evaluating asymptotic p-values can be numerically unstable, Monte-Carlo methods can be used to verify interesting p-values.

Results

We first evaluated our method using simulations. Genes were simulated as stretches of 5,000 base-pairs using the coalescent[16] and a demographic model (including an ancient bottleneck, recent exponential growth, differentiation and migration) calibrated to mimic a sample of multiple European populations[17,18] (Supplementary Figure 1 and Supplementary Notes). The average F_{ST} value between simulated populations was 0.004 – as expected when the distribution of rare variants is geographically restricted[19]. The simulations produced samples of 1,000 individuals, each drawn from one of several related populations, typically including a few shared variants and many population specific variants. Half of the simulated variants were randomly set to increase trait values by $1/8^{th}$ of a standard deviation (Supplementary Figure 2 and see Supplementary Figure 3 and 4 for similar results using alternative trait models).

We analyzed each simulated sample with a series of gene-level association tests. Supplementary Figures 2-4 compare results obtained for 10,000 simulated genes using our meta-analysis approach to a combined analysis of individual level data across studies. For variable threshold tests, we found the p-values were sometimes slightly different (r^2=0.995 between the two sets of log p-values); for the other two tests p-values and test statistics were indistinguishable. Calculation of analytical p-values for variable threshold tests requires the evaluation of high-dimensional integrals that can be numerically unstable and is thus very sensitive to small differences in the variance-covariance matrix. In practice, it will often be a good idea to confirm significant p-values using our Monte-Carlo approach.

To evaluate our Monte-Carlo approach, we compared its empirical p-values to those obtained by permuting phenotypes between individuals within each study. We implemented adaptive versions of both algorithms[20], with more simulations carried out when the p-value

is small and fewer simulations when the p-value is large. Log p-values for the two approaches are highly concordant (r^2=0.996). When small p-values are estimated, increasing the number of simulations improves the precision for the estimated p-values (Supplementary Figure 5).

We next verified type I error was well controlled (Supplementary Table 1). In all analyses, we first applied an inverse normal transformation to trait residuals (which helps ensure our statistics are well behaved even for very rare variants, as in Supplementary Figure 6). Reassured that type I error was well controlled, we next explored power for several scenarios (Figure 1A, 1B, 1C and Supplementary Figure 7A, 7B, 7C). It is clear that, for the effect sizes simulated here, very large samples may be required. In some settings, power only reaches ∼60% in analyses of ∼100,000 individuals. We did not find a universally most powerful method, emphasizing the value of implementing a diverse set of test statistics (see also Ladouceur et al[13]). Since meta-analysis methods that combine p-values are popular for common variants and can also be implemented for rare variants, we compared power between our method and analyses based on Fisher's method and the minimal p-value approach for combining p-values (Figure 1 **and** Supplementary Figure 7). In all the simulation scenarios considered, our method greatly outperforms these alternatives, especially when information is combined across a large number of samples. In addition to power, our approach provides three useful features. First, it provides great flexibility in the choice of rare variant association test (definition of functional units, choice of variants to be grouped, frequency thresholds for analysis); approaches based on Fisher's method would likely require every contributing study to re-analyze their data when any of these changes. Second, because in addition to p-values it provides for estimates of effect size (in all cases) and allele frequency thresholds for candidate variants (in the variable threshold test), our method provides rich information that helps interpretation. Third, our approach allows the relationship between multiple association signals in a region to be dissected through conditional analysis, as detailed below.

We proceeded to a meta-analysis of blood lipid levels in 18,699 individuals of European ancestry genotyped with Illumina Exome arrays and drawn from 7 studies: the Women's Health Initiative[21], the Ottawa Heart Study[22], the Malmö Diet and Cancer Study – Cardiovascular Cohort (MDC)[23], the PROCARDIS Precocious Coronary Artery Disease Case Series, PROCARDIS Control series[24] and the Nord-Trøndelag Health Study (HUNT) myocardial infraction cases and matched controls[25] (see Supplementary Table 2 and 3 for summary statistics for each of these samples, including basic demographics, summaries of lipid levels, number of non-synonymous and loss-of-function variants per individual and of variants sites shared across different studies). Overall, 171,193 variants were polymorphic in at least one individual. Among these variants, 125,702 – the vast majority – have frequency <1%.

To verify the soundness of our approach, we repeated our power and type I error simulations using real genotype data from the HUNT and MDC studies but simulated phenotypes. These additional experiments confirm that our method produces well-calibrated statistics and is more robust to stratification than analyses that directly pool individual level data and treat the complete dataset as a single study without modeling heterogeneity between studies (Supplementary Figure 8). In addition, the power for our method continued to exceed that for alternatives that directly combine p-values from individual studies (Supplementary Figure 9).

We then proceeded to meta-analyze single variant association test results. The resulting test statistics appear well calibrated, with genomic control value <1.05 for all three traits, both for common and for rare variants (Supplementary Figure 10). At a significance threshold of

$p<3\times10^{-7}$ (corresponding to 0.05/171,193), we found significantly associated variants (with MAF<5%) at *LPL*[26], *ANGPTL4*[26], *LIPG*[26], *CD300LG*[27], *LIPC*[26], *APOB*[26], *HNF4A*[26] for HDL; *PCSK9*[26], *BCAM-CBLC-PVR* (neighboring APOE)[26], and *APOB*[26] for LDL; *ANGPTL4*[26], *LPL*[26] and *APOB*[26] for TG (Supplementary Table 4). Except for the variants in *LIPC* and *APOB*, all other significantly associated variants have frequency of >1% reflecting the limited power of single variant association tests for rare alleles.

We next carried out gene-level tests. Again, test statistics appear well calibrated, with genomic control value <1.05 (Supplementary Figure 11). At a significance threshold of $p<3.1\times10^{-6}$ (corresponding to 0.05/16,153 and thus allowing for the number of genes tested), we observed association at *LIPC*, *LPL*, *ANGPTL4*, *LIPG*, *HNF4A* and *CD300LG* for HDL, at the *PCSK9*, *APOE*-locus (as well as nearby genes *PVR*, *BCAM*, and *CBLC*), and *LDLR* for LDL, and at *ANGPTL4*, and *LPL* for triglycerides (Table 1). Supplementary Table 5 emphasizes that, at these loci, much stronger signals are identified in meta-analysis than in any component study. Reassuringly, these signals point to loci identified in previous genome-wide association studies and/or re-sequencing studies. Importantly, note that our approach was able to appropriately identify the signal in *LDLR* which is driven by several very rare variants (each with frequency < .00052) that nearly always increase blood LDL cholesterol levels and that, at several other loci, gene-level p-values exceeded the best single variant p-value in the gene (Supplementary Table 6). We again compared our method with conventional methods such as minimal p-value approach, Fisher's method, and an extended Fisher's method taking into account unequal sample sizes (**Methods**). As shown in Supplementary Tables 7-9, our method identifies a larger number of loci, all known to be associated with lipid levels in humans. We also compared results obtained from our meta-analysis method with results from directly pooling a subset of the data (after normal transformation of trait values in each sample to avoid artifacts due to stratification). Reassuringly, p-values from our approach and joint analysis of pooled data were highly concordant with $r^2>0.99$ (Supplementary Figure 12), in accordance with results obtained using coalescent simulations.

An added convenience of sharing single-variant statistics together with their covariance matrices, as we propose, is that it facilitates conditional analyses, extending an idea used by Yang et al[28] for analysis of common variants in GWAS meta-analysis. Supplementary Figure 13 illustrates how, in simulations, common variants can generate shadow rare variant association signals at nearby genes, and how our method for conditional analysis resolves the problem. In real data, we re-examined two of the LDL associated loci in detail, *LDLR* and *APOE-BCAM-CBLC-PVR*. For *LDLR*, we examined the relationship between rare variant signals and three nearby common variants[26]. Specifically, we conditioned on genotypes for 3 common variants (rs6511720, rs2228671 and rs72658855) exhibiting significant association in the region, and found that *LDLR* rare variant association remains significant (p-value 4.6×10^{-7}) (Supplementary Table 10). For the *APOE-BCAM-CBLC-PVR* locus, after conditioning on the common variant showing strongest association in the region (rs7412), gene-level associations at *BCAM*, *CLBC* and *PVR* become non-significant, suggesting that these rare-variant signals are the result of regional linkage disequilibrium with more common and well described variants in *APOE* (Supplementary Table 11). We also analyzed top single association signal conditional on the genotypes of rare variants (with MAF≤5%) that are included in the burden tests. We showed that the top single variant signals from both *APOE* gene and the *LDLR* gene remained significant (Supplementary Table 12). For completeness, Supplementary Figure 14 and 15 show that conditional analyses using individual level data in a subset of samples and conditional analyses using our meta-analysis based approach give highly concordant p-values ($r^2>0.99$).

Discussion

In the analysis of each sample, when population stratification is of concern, we recommend that principal components of genotype matrix should be incorporated in the regression model as covariates[29] or that linear mixed models with empirically estimated kinship matrices should be used[30]. Linear mixed models can also be used to account for relatedness in family studies or other samples that include cryptically related individuals. Our software implementation readily allows for both these options, including correct calculation of kinship matrices to allow family samples to be included in meta-analyses (see **Methods** for details).

Although we only presented applications of our method to quantitative trait meta-analysis, our methods and tools can be applied to binary traits as well (see **Methods** for details). For binary traits, distributions about normality of test statistics may be less reliable. These could affect performance of our resampling method for empirical p-values, meta-analysis results for the rarest variants, and conditional analysis statistics (see also the work of Lin and Tang [9] and Lee et al [31]). Since performance of our methods (and other similar approaches) for binary traits will depend on factors like sample size and the balance of cases and controls in each sample, we recommend careful quality control of results for such studies, including for example, review of quantile-quantile plots for variants of different frequencies. Our methods are implemented as freely available software, including programs for calculating summary statistics, annotating the resulting summaries, performing meta-analysis, calculating gene-level statistics and executing conditional analyses. Our tools work with standard VCF files[32] for genotype data and Merlin[33] or PLINK[34] files for phenotype data.

Meta-analysis has facilitated many discoveries in common variant association studies. Here, we describe a powerful framework for meta-analysis of rare variants at the level of genes or other functional units. Through simulation and empirical evaluation, we demonstrate that our approach is well calibrated and provides comparable power to more cumbersome analyses that require pooling all individual level data. Through the analysis of blood lipids levels across seven studies, we show that our approach can detect rare variant association signals at known candidate loci. Our method has a variety of unique features, which include supporting a variety of rare variant association tests, allowing for the analysis of family samples and the calculation of empirical p-values, and for conditional analysis that can distinguish truly novel rare variant signals from shadows of other nearby common or rare associations. We envision that this approach (and continued development of related approaches [35-37]) will facilitate the large sample sizes required to accelerate new discoveries in complex trait genetics.

Methods

This section starts with a summary of notation, proceeds to describe the statistics to be shared between studies and methods for single variant meta-analysis. We then show that the statistics for different gene-level tests can be calculated using summary level data, enabling efficient meta-analysis. In the Supplementary Notes, we provide many additional details and summarize how each of the test statistics used here can be derived as a score test using likelihood functions that allow for per-sample nuisance parameters.

Notation

For simplicity, we describe our strategy for analysis of a single gene. Let J be number of variant nucleotide sites genotyped in at least one study. For study k, let n_k denote the number of samples phenotyped and genotyped, and let the vector $\mathbf{y_k} = (Y_{1,k},\ldots,Y_{N_k,k})^T$ denote the quantitative trait residuals (after adjustment for any covariates), with variance σ_k^2. Within

each study *k*, we encode genotype information in matrix $\boldsymbol{X_k}$ where each entry $X_{i,j,k}$ represents the genotype for individual *i* at site *j*, coded as the number of alternative alleles. We encode missing genotypes in the dataset as the average number of minor alleles in individuals who are genotyped for that marker. The multi-site genotype for individual *i* is denoted by the row vector $\mathbf{x}_{i,\bullet,k}$, and the genotypes for all N_k individuals at site *j* are given by column vector $\mathbf{x}_{\bullet,j,k}$ For the ease of presentation, we define the mean genotype matrix $\bar{\boldsymbol{X}}_{\boldsymbol{k}}$, where the (*i*,*j*)-th element is $(\Sigma_i X_{i,j,k})/N_k$.

Summary Statistics To Be Shared

For each study, we first calculate and share a vector of score statistics $\mathbf{u_k} = (\boldsymbol{X_k} - \bar{\boldsymbol{X}}_{\boldsymbol{k}})^T \mathbf{y_k}$, a corresponding variance-covariance matrix

$\boldsymbol{V_k} = \hat{\sigma}_k^2 N_k \mathrm{cov}(\boldsymbol{X_k}) = \hat{\sigma}_k^2 \left(\boldsymbol{X_k} - \overline{\boldsymbol{X}}_{\boldsymbol{k}}\right)^T \left(\boldsymbol{X_k} - \overline{\boldsymbol{X}}_{\boldsymbol{k}}\right)$, and allele frequencies for each marker $p_{j,k} = \Sigma_i X_{i,j,k}/2N_k$. Note that $\boldsymbol{V_k}$ effectively describes linkage disequilibrium relationships between the variants being examined. To perform quality control, we also share mean and variance for the quantitative trait residuals, genotype call rate and Hardy-Weinberg equilibrium p-values at each variant site.

Meta-analysis of Single Variant Association Test Statistics

We first combine single variant association test statistics across studies using the Cochran-Mantel-Haenszel method. Specifically, we calculate a score statistic at each site as:

$$t_{j,\bullet} = U_{j,\bullet} / \sqrt{V_{j,j,\bullet}} \quad (1)$$

where $U_{j,\bullet} = \Sigma_k U_{j,k}$ and $V_{j,j,\bullet} = \Sigma_k V_{j,j,k}$. For ease of presentation, we denote the vector of single variant association tests after meta-analysis as $\mathbf{u} = \Sigma_k \mathbf{u_k}$. Under the null, this vector is distributed as multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\Sigma_k \mathbf{V_k}$.

Burden Tests That Assume Variants Have Similar Effect Sizes

For a simple burden test in study *k*, the impact of multiple rare variants in a region can be modeled using a shared regression coefficient in a model that takes the form:

$$Y_{i,k} = \beta_{0,k} + \beta_{BURDEN} C_{BURDEN}(\mathbf{x_{i,\bullet,k}}) + \varepsilon_{i,k}, \text{where } \varepsilon_{i,k} \sim \mathrm{N}\left(0, \sigma_k^2\right) \quad (2)$$

$C_{BURDEN}(\mathbf{x_{i,\bullet,k}})$ is a function that takes genotypes for a single individual as input and returns the count of rare alleles (the "rare variant burden") in the gene being examined. When individual level data is available and nuisance parameters $\beta_{0,k}$ and σ_k^2 are allowed to vary between studies, the score statistic for a rare variant burden test becomes:

$$U_{BURDEN} = \sum_k U_{BURDEN,k} = \sum_k \boldsymbol{\omega}^{\mathbf{T}} \mathbf{u_k} = \boldsymbol{\omega}^{\mathbf{T}} \mathbf{u} \quad (3)$$

which is equal to a linear sum of (weighted) single variant score statistics.

Under the null, this statistic is approximately normally distributed with mean 0 and variance $V_{BURDEN} = \boldsymbol{\omega}_T (\Sigma_k \boldsymbol{V_k}) \boldsymbol{\omega}$, enabling significance tests. Here, $\boldsymbol{\omega}$ is the vector of weights, which is $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_j)$, with each element ω_j representing the weight assigned to variant *j* according to its allele frequency or its computationally predicted functional impact[10,15]. The formula above makes it clear that, when nuisance parameters are allowed to vary between

studies, the same burden score statistics that could be calculated by sharing individual data can be equivalently calculated using shared summary statistics.

Variable Threshold Tests with an Adaptive Frequency Threshold

In variable threshold test, rare variant burden statistics are calculated for each observed variant minor allele frequency threshold and significance is evaluated for the maximum of these statistics. Given a specific variant frequency threshold *F* we define the resulting burden score statistic as:

$$U_{BURDEN(F)} = \mathbf{v_F^T}\overrightarrow{U}. \quad (4)$$

Here, $\mathbf{v_F}$ is a vector of indicators where the j^{th} element equals 1 if the pooled minor allele frequency at variant site *j* is less than *F* and zero otherwise. For convenience, we also define a matrix of indicators for minor allele frequency thresholds $\mathbf{\Phi} = (\mathbf{v_{F_1}}, \mathbf{v_{F_2}}, \ldots, \mathbf{v_{F_J}})$. After a burden statistic is calculated for each potential frequency threshold, these are standardized, dividing each statistic by its corresponding variance, and the maximum statistic is identified:

$$T_{VT} = \max_F \left\{T_{BURDEN(F)}\right\}, \text{where } T_{BURDEN(F)} = U_{BURDEN(F)} / \sqrt{\mathbf{v}_F^T \sum_k \boldsymbol{V_k} \mathbf{v}_F} \quad (5)$$

Significance for this statistic can be evaluated using the cumulative distribution function for the multivariate normal distribution[38]. Specifically, given the definition of the covariance between burden statistics calculated using different allele frequency thresholds, we have:

$$\left(T_{BURDEN(F_1)}, \cdots, T_{BURDEN(F_M)}\right) \sim \text{MVN}\left(\mathbf{0}, \boldsymbol{\Phi}\left(\sum_k \boldsymbol{V_k}\right)\boldsymbol{\Phi^T}\right) \quad (6)$$

The p-value for the VT test statistic is given by

$$p = 1 - \Pr\left(T_{VT} \le t_{VT}\right) = 1 - \Pr\left(T_{BURDEN(F_1)} \le t, \cdots, T_{BURDEN\left(F_M\right)} \le t\right) = 1 - F_{MVN}(t, \cdots, t), \quad (7)$$

where F_{MVN} is the distribution function for the multivariate normal distribution $\text{MVN}(\mathbf{0}, \boldsymbol{\Phi}(\sum_k \boldsymbol{V_k})\boldsymbol{\Phi^T})$.

Burden Tests that Assume A Distribution of Variant Effect Sizes (e.g. SKAT tests)

The simple burden test and variable threshold test described above can be underpowered when variants with opposite phenotypic effects reside in the same gene and are grouped together, because the shared regression coefficient can average close to zero in that situation[9-12]. To accommodate this setting, we consider an underlying distribution of rare variance effect sizes with mean zero and test whether the variance of this distribution τ is greater than zero.

When individual level data is available, association analysis in study *k* is performed using the following model

$$Y_{i,k} = \beta_{0,k} + \sum_j \beta_j X_{i,j,k} + \varepsilon_{i,k}, \text{where } \varepsilon_{i,k} \sim \text{N}\left(0, \sigma_k^2\right) \quad (8)$$

We make inferences about rare variant effect sizes $\beta = (\beta_1, \beta_2, \ldots \beta_J)$ by assuming these follow a common distribution with mean zero and variance τ. Under the null, $\tau=0$. Following Wu et al[9], in Supplementary Notes we derive the score statistic for this model and show that it can be calculated on the basis of per-study summary statistics:

$$Q=\left(\sum_k \mathbf{u_k}\right)^T \boldsymbol{K} \left(\sum_k \mathbf{u_k}\right) \quad (9)$$

Here, $\boldsymbol{K}$ is the kernel matrix that compares multi-site genotypes. A default choice[9] is a diagonal matrix $\boldsymbol{K} = diag(\omega_1, \omega_2 \ldots, \omega_J)$, with ω_j being the weight assigned to variant site j. The statistic Q follows a mixture chi-square distribution[31], which means that Q is equivalent in distribution to a weighted sum of independent chi-square random variables. The weights (or mixture proportions) are given by the eigenvalues for the matrix $(\Sigma_k \boldsymbol{V_k})^{1/2}\, \boldsymbol{K} (\Sigma_k \boldsymbol{V_k})^{1/2}$.

Monte-Carlo Method for Empirical Assessment of Significance

The previous sections describe how a series of gene-level test statistics can be calculated and, for each one, propose a strategy for evaluating significance using asymptotic distributions. In practice, evaluating the required numerical integrals can be challenging because variance-covariance matrices that are sometimes singular or nearly singular.

Note that single variant test statistics are distributed as:

$$\sum_k \mathbf{u_k} = \sum_k \mathbf{y_k^T} \left(\boldsymbol{X_k} - \overline{\boldsymbol{X}}_{\boldsymbol{k}}\right) \sim \mathrm{MVN}\left(\mathbf{0}, \sum_k \boldsymbol{V_k}\right) \quad (10)$$

Then, to evaluate significance empirically, one can sample random vectors from the distribution MVN($\mathbf{0}$, $\Sigma_k \boldsymbol{V_k}$) and calculate gene-level rare variant test statistics for each of these sampled random vectors, resulting in an empirical distribution for any gene-level statistic[39]. As usual, p-values can then be evaluated by comparing the test statistics for the original data with those in this empirical distribution. For computational efficiency, we use an adaptive algorithm where a larger number of vectors are sampled when assessing small p-values and fewer vectors are sampled when assessing larger p-values[20].

Conditional Analyses

It is well known that, due to linkage disequilibrium, one or more common causal variants can result in shadow association signals at other nearby common variants. For common variants, Yang et al[28] have shown that linkage disequilibrium relationships between variants, estimated from external reference panels, can be used to enable conditional analysis in meta-analysis settings. For rare variants and gene-level tests, accurately describing relationships between variants is crucial and we recommend against the use of external reference panels. Instead, in the Supplementary Notes, we describe how conditional analysis statistics can be derived for different gene-level tests in our meta-analysis setting.

Analysis of Samples of Known or Hidden Relatedness

Our methods and tools can also be used when samples within a study are related to each other. Detailed formulae of the score statistics and their covariance matrices when linear mixed models are used to account for relatedness, are described in the Supplementary Notes.

Analysis of Dichotomous Trait

Our approach extends naturally to the analysis of binary traits. Specifically, when single variant score statistics and their covariance matrices are shared, meta-analysis test statistics

can be calculated in the same manner as for continuous trait. Detailed definitions of test statistics for binary traits are given in the Supplementary Notes. A limitation is that, when variant counts in a gene or analysis unit are very small or the number of cases and controls in each study is very unbalanced, the asymptotic distributions for burden statistics may not hold, and p-values obtained using our approach may not be accurate. In practice, we recommend careful review of QQ plots for meta-analysis statistics (as is standard in genome-wide association studies).

Weighted Fisher's Methods, Incorporating Unequal Sample Sizes

To accommodate the scenario where samples of different sizes are meta-analyzed, we use a modified version of Fisher's method that incorporates sample sizes as weights for each study. Specifically, our test statistic is defined by $T_{Weighted–Fisher} = -2\Sigma_k N_k \log p_k$. The weighted Fisher's test statistic follows a mixture chi-square distribution with mixture proportions given by $N_1, N_1, N_2, N_2, \ldots, N_k, N_k$.

Simulation of Population Genetic Data

We simulated haplotypes using a coalescent model and the program *ms*[16]. We chose a demographic model consistent with European demographic history[4], including an ancestral bottleneck followed by more recent population differentiation and exponential growth. Model parameters were based upon estimates from large scale sequencing studies[40], as detailed in Supplementary Notes.

Meta-Analysis of Lipid Traits

Summary statistics were calculated for each participating study and shared to enable a central meta-analysis. In single variant and gene-base rare variant association analysis, age, age^2, sex and cohort specific covariates, such as principal components of ancestry were included in the analysis. Trait residuals were standardized using inverse normal transformation. More detailed descriptions for each participating cohort are given in the Supplementary Notes. This research was approved by the Institutional Review Board of the University of Michigan and the Broad Institute. Informed consent was obtained from all study subjects. In addition, all participating studies received approvals from local ethics committee.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Drs. Michael Boehnke, Xiaoquan (William) Wen, and Sebastian Zoellner for helpful discussions. This work was supported by research grants R01HG007022 from the National Human Genome Research Institute, R01EY022005 from the National Eye Institute and R01HL117626 from the National Heart, Lung and Blood Institute. G.M.P was supported by Award Number T32HL007208 from the National Heart, Lung, and Blood Institute. S.K. is supported by a Research Scholar award from the Massachusetts General Hospital (MGH), the Howard Goodman Fellowship from MGH, the Donovan Family Foundation, and R01HL107816. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. This manuscript was prepared in collaboration with investigators of the WHI, and has been approved by the WHI. WHI investigators are listed at http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf. For full list of PROCARDIS acknowledgements, visit www.procardis.org. The Ottawa Heart Genomics Study was supported by the Canadian Institutes of Health Research (CIHR) MOP-82810, MOP-77682, MOP-2380941 and the Canada Foundation for Innovation (CFI) 11966. The studies in Malmö Diet and Cancer cohort were supported by grants from the Swedish Research Council, the Swedish Heart and Lung

Foundation, the Påhlsson Foundation, the Novo Nordic Foundation and an European Research Council Starting grant StG-282255.

References

1. 1000 Genomes Project, C. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
2. Kiezun A, et al. Exome sequencing and the genetic basis of complex traits. Nat Genet. 2012; 44:623–30. [PubMed: 22641211]
3. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet. 2008; 83:311–21. [PubMed: 18691683]
4. Kryukov GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR. Power of deep, all-exon resequencing for discovery of human trait genes. Proc Natl Acad Sci U S A. 2009; 106:3871–6. [PubMed: 19202052]
5. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2009
6. Price AL, et al. Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet. 2010; 86:832–8. [PubMed: 20471002]
7. Liu DJ, Leal SM. Replication strategies for rare variant complex trait association studies via next-generation sequencing. Am J Hum Genet. 2010; 87:790–801. [PubMed: 21129725]
8. Zawistowski M, et al. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. Am J Hum Genet. 2010; 87:604–17. [PubMed: 21070896]
9. Wu MC, et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011; 89:82–93. [PubMed: 21737059]
10. Lin DY, Tang ZZ. A general framework for detecting disease associations with rare variants in sequencing studies. Am J Hum Genet. 2011; 89:354–67. [PubMed: 21885029]
11. Lin X. Variance component testing in generalised linear models with random effects. Biometrika. 1997; 84:309–326.
12. Neale BM, et al. Testing for an unusual distribution of rare variants. PLoS Genet. 2011; 7:e1001322. [PubMed: 21408211]
13. Ladouceur M, Dastani Z, Aulchenko YS, Greenwood CM, Richards JB. The empirical power of rare variant association methods: results from sanger sequencing in 1,998 individuals. PLoS Genet. 2012; 8:e1002496. [PubMed: 22319458]
14. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2010; 34:188–93. [PubMed: 19810025]
15. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet. 2009; 5:e1000384. [PubMed: 19214210]
16. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 2002; 18:337–8. [PubMed: 11847089]
17. Adams AM, Hudson RR. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. Genetics. 2004; 168:1699–712. [PubMed: 15579718]
18. Novembre J, et al. Genes mirror geography within Europe. Nature. 2008; 456:98–101. [PubMed: 18758442]
19. Nelson MR, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 2012; 337:100–4. [PubMed: 22604722]
20. Besag J, Clifford P. Sequential Monte Carlo p-values. Biometrika. 1991; 78:301–304.
21. Tennessen JA, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012; 337:64–9. [PubMed: 22604720]
22. McPherson R, et al. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007; 316:1488–91. [PubMed: 17478681]
23. Kathiresan S, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med. 2008; 358:1240–9. [PubMed: 18354102]

24. Clarke R, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N Engl J Med. 2009; 361:2518–28. [PubMed: 20032323]
25. Krokstad S, et al. Cohort Profile: The HUNT Study, Norway. Int J Epidemiol. 2012
26. Teslovich TM, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–13. [PubMed: 20686565]
27. Albrechtsen A, et al. Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. Diabetologia. 2013; 56:298–310. [PubMed: 23160641]
28. Yang J, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012; 44:369–75. S1–3. [PubMed: 22426310]
29. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
30. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012; 44:821–4. [PubMed: 22706312]
31. Lee S, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012; 91:224–37. [PubMed: 22863193]
32. Danecek P, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156–8. [PubMed: 21653522]
33. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002; 30:97–101. [PubMed: 11731797]
34. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]
35. Hu YJ, et al. Meta-analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics. Am J Hum Genet. 2013
36. Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. Am J Hum Genet. 2013; 93:42–53. [PubMed: 23768515]
37. Tang ZZ, Lin DY. MASS: meta-analysis of score statistics for sequencing studies. Bioinformatics. 2013; 29:1803–5. [PubMed: 23698861]
38. Genz A. Numerical Computation of Multivariate Normal Probabilities. Journal of Computational and Graphical Statistics. 1992; 1:141–149.
39. Zou F, Fine JP, Hu J, Lin DY. An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative trait Loci. Genetics. 2004; 168:2307–16. [PubMed: 15611194]
40. Coventry A, et al. Deep resequencing reveals excess rare recent variants consistent with explosive population growth. Nat Commun. 2010; 1:131. [PubMed: 21119644]

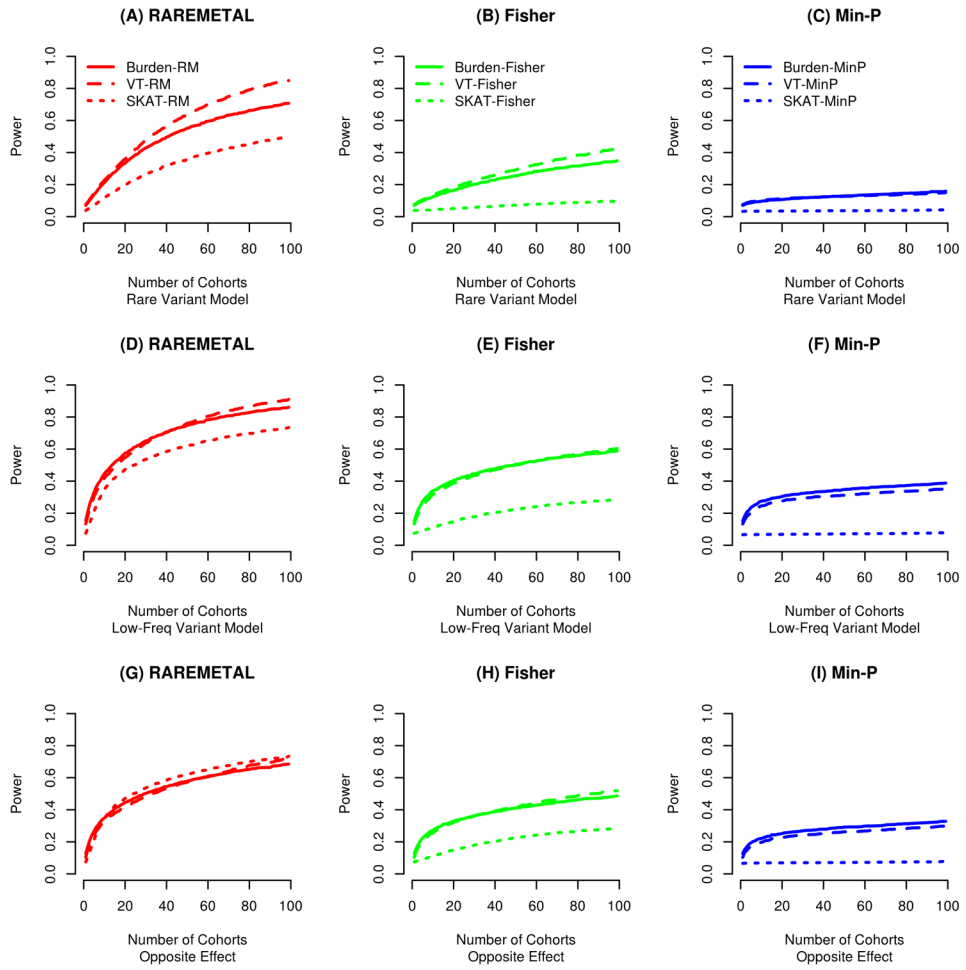

Figure 1.

Power comparison for our approach, Fisher's method and the minimal p-value approach. Three phenotype models were simulated: (1) half of low frequency variants with MAF < 0.5% are causal, each increasing expected trait values by 1/4 standard deviation; (2) half of all variants are causal, irrespective of frequency, and increase trait values by 1/4 standard deviation; (3) 50% of the variants are casual, irrespective of frequency, and 80% of these increase expected trait values by 1/4 standard deviation, while the remaining 20% decrease trait values by the same amount. A number of 2-100 samples of size 1000 were simulated for each model, with each sample drawn from a randomly chosen population. Meta-analysis was performed using our approach or using Fisher's method and the minimal p-value approach to combine burden test, SKAT and variable threshold (VT) test statistics for variants with MAF<5%. The power was evaluated at the significance threshold of $\alpha=2.5\times10^{-6}$ using 10,000 replicates. Panel A displays the power for three meta-analysis methods using simple burden test under model (1). Panel B displays the results for three meta-analysis methods using VT under model (1). Panel C displays the results for three meta-analysis methods using SKAT under model (1). Panel D displays the results for three meta-analysis methods using simple burden test under model (2). Panel E displays the results for three meta-analysis methods using VT under model (2). Panel F displays the results for three meta-analysis methods using SKAT under model (2). Panel G displays the results for three meta-analysis methods using simple burden test under model (3). Panel H displays the results for three meta-analysis methods using VT under model (3). Panel I displays the results for three meta-analysis methods using SKAT under model (3). Note that

differences between our approach and these alternatives become more marked when more studies are meta-analyzed.

Table 1

Results for meta-analysis of gene-level rare variant association test. Associations that attain exome-wide significance ($p < 3.1\times10^{-6}$) are displayed. Five gene-level association tests were used to analyze the data: simple burden tests with 1% or 5% cutoff (Burden-1 and Burden-5), SKAT tests with 1% or 5% cutoff (SKAT-1 and SKAT-5) and variable threshold (VT) tests that analyze variants with MAF<5%. Significant p-values for each test are displayed in bold font. For the associations that are significant, estimates of average genetic effect are also shown. The loci where one or more gene-based association signal exceeds the top single variant association signal are labeled with an asterisk.

Gene	Gene Position[a]	Burden-1	Burden-5	SKAT-1	SKAT-5	VT	MAF Cutoff	Direction of Single Variant Association Statistics[b]	Estimates of Genetic Average Effect (s.d units) for Rare Variants under Different MAF Thresholds		
									0.01	0.05	VT
HDL											
LIPC*	chr15:58.7Mb	**1.4×10^{-12}**	**3.5×10^{-7}**	**1.8×10^{-9}**	1.4×10^{-2}	**4.5×10^{-12}**	3.7×10^{-3}	-++++--+-	0.5	0.1	0.5
LPL*	chr8:19.8Mb	9.7×10^{-1}	**2.5×10^{-24}**	3.5×10^{-1}	**5.0×10^{-13}**	**1.5×10^{-23}**	2.5×10^{-2}	(-)-(-)+-++	-	-0.3	-0.3
ANGPTL4*	chr19:8.4Mb	2.2×10^{-2}	**2.9×10^{-19}**	2.2×10^{-2}	**3.0×10^{-19}**	**1.8×10^{-18}**	2.6×10^{-2}	(+)--++-+++	-	0.3	0.3
LIPG*	chr18:47.1Mb	2.2×10^{-5}	**6.4×10^{-19}**	2.1×10^{-5}	**2.9×10^{-9}**	**4.4×10^{-18}**	1.3×10^{-2}	-++----(+)+	-	0.4	0.4
HNF4A	chr20:43.0Mb	7.5×10^{-1}	**2.8×10^{-7}**	6.8×10^{-1}	**2.5×10^{-7}**	**1.5×10^{-6}**	4.1×10^{-2}	(-)--+-+	-	-0.1	-0.1
CD300LG	chr17:41.9Mb	4.9×10^{-1}	**8.5×10^{-7}**	5.2×10^{-1}	1.0×10^{-5}	**3.1×10^{-6}**	3.3×10^{-2}	(-)+-(+)	-	-0.1	-
LDL											
PCSK9*	chr1:55.5Mb	1.8×10^{-2}	**7.4×10^{-19}**	8.1×10^{-2}	**5.5×10^{-17}**	**2.0×10^{-28}**	1.3×10^{-2}	(-)--(-)--+-++-	-	-0.3	-0.5
BCAM	chr19:45.3Mb	1.7×10^{-1}	**1.6×10^{-18}**	1.5×10^{-1}	3.0×10^{-5}	**2.6×10^{-17}**	3.6×10^{-2}	(-)+++(-)+-+++---+(-)+--+--++	-	-0.1	-0.1
CBLC	chr19:45.3Mb	9.4×10^{-1}	**2.0×10^{-15}**	4.4×10^{-1}	1.5×10^{-4}	**1.0×10^{-14}**	4.4×10^{-2}	-(-)--+-(-)(+)	-	-0.1	-0.1
PVR	chr19:45.2Mb	6.1×10^{-2}	**3.0×10^{-10}**	4.8×10^{-2}	6.3×10^{-2}	**1.1×10^{-9}**	4.9×10^{-2}	(-)++--+	-	-0.1	-0.1
LDLR*	chr19:11.2Mb	1.8×10^{-3}	4.7×10^{-5}	3.8×10^{-2}	2.5×10^{-1}	**2.4×10^{-7}**	5.2×10^{-4}	+++++++++-++++--+	-	-	0.8
TG											
ANGPTL4*	chr19:8.4Mb	2.6×10^{-2}	**1.2×10^{-24}**	3.7×10^{-2}	**3.9×10^{-25}**	**7.1×10^{-24}**	2.6×10^{-2}	(-)+---+---	-	-0.3	-0.2
LPL*	chr8:19.8Mb	6.8×10^{-1}	**7.7×10^{-20}**	2.6×10^{-1}	**1.8×10^{-11}**	**4.6×10^{-19}**	2.5×10^{-2}	(+)+(+)--+-	-	0.2	0.2

[a] Gene position is defined based upon hg19, GRCh37 Genome Reference Consortium Human Reference 37

[b] Direction of single site statistics for variants with MAF<5%. Variants within parenthesis have frequency >1%.

* The loci with one or more gene-level association signal exceeding the top single variant signal.