Methods for Association Analysis and Meta-Analysis of Rare Variants in Families

Shuang Feng,[1] Giorgio Pistis,[1,2] He Zhang,[3] Matthew Zawistowski,[1] Antonella Mulas,[2] Magdalena Zoledziewska,[2] Oddgeir L. Holmen,[4,5] Fabio Busonero,[2] Serena Sanna,[2] Kristian Hveem,[4,6] Cristen Willer,[3,7,8] Francesco Cucca,[2] Dajiang J. Liu,[9] and Gonçalo R. Abecasis[1]*

[1]Department of Biostatistics, Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, United States of America; [2]Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, Monserrato, Italy; [3]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, United States of America; [4]Department of Public Health and General Practice, HUNT Research Centre, Norwegian University of Science and Technology, Levanger, Norway; [5]St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; [6]Department of Medicine, Levanger Hospital, Nord-Trøndelag Health Trust, Levanger, Norway; [7]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America; [8]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, United States of America; [9]Department of Public Health Sciences, College of Medicine, Penn State University Hershey, Pennsylvania, United States of America

ABSTRACT: Advances in exome sequencing and the development of exome genotyping arrays are enabling explorations of association between rare coding variants and complex traits. To ensure power for these rare variant analyses, a variety of association tests that group variants by gene or functional unit have been proposed. Here, we extend these tests to family-based studies. We develop family-based burden tests, variable frequency threshold tests and sequence kernel association tests. Through simulations, we compare the performance of different tests. We describe situations where family-based studies provide greater power than studies of unrelated individuals to detect rare variants associated with moderate to large changes in trait values. Broadly speaking, we find that when sample sizes are limited and only a modest fraction of all trait-associated variants can be identified, family samples are more powerful. Finally, we illustrate our approach by analyzing the relationship between coding variants and levels of high-density lipoprotein (HDL) cholesterol in 11,556 individuals from the HUNT and SardiNIA studies, demonstrating association for coding variants in the *APOC3, CETP, LIPC, LIPG,* and *LPL* genes and illustrating the value of family samples, meta-analysis, and gene-level tests. Our methods are implemented in freely available C++ code.

Genet Epidemiol 39:227–238, 2015. © 2015 Wiley Periodicals, Inc.

KEY WORDS: Rare variant; association; family sample; population sample; gene-level tests; study design; meta-analysis

Introduction

Variants of functional consequence, including nonsynonymous, splice altering, and protein truncating variants, usually segregate at very low frequency in human populations [Abecasis et al., 2010, 2012; Marth et al., 2011; Nelson et al., 2012]. Recent advances in exome sequencing and the development of exome genotyping arrays are enabling explorations of their contributions to complex disease [Kiezun et al., 2012].

Association of rare variants with disease will bring biological insights about disease processes, but standard variant-by-variant association tests lack power when applied to these variants. Our work builds upon three strategies to increase the power of rare variant association studies: grouping variants by gene or functional unit, combining results across many studies through meta-analysis, and analysis of family samples.

Grouping rare variants by gene or functional unit [Li and Leal, 2008], whether with weights [Madsen and Browning, 2009] or without [Morris and Zeggini, 2010], is now a popular strategy for rare variant association analysis [Lee et al., 2012a,b; Lin and Tang, 2011; Price et al., 2010; Wu et al., 2011]. The approach assumes that rare variants in the same gene or functional unit have similar functional consequences. When the assumption is correct and rare variants in a region are analyzed together, association signals will be clearer than when evaluating variants individually.

A second strategy to increase power is meta-analysis, which increases sample size and provides a practical approach to difficulties in data-sharing and concerns about heterogeneity [Lin and Zeng, 2010; Willer et al., 2010]. Meta-analysis of single variants has been key in establishing association between common variants and complex diseases [Scott et al., 2007; Willer et al., 2010]. Meta-analysis methods for rare variant association tests have now been proposed, although these initial proposals and their implementations have generally

focused on samples of unrelated individuals [Lee et al., 2013; Liu et al., 2014; Tang and Lin, 2013].

Finally, a third strategy is to study samples of closely related individuals, increasing the odds that multiple copies of each rare variant are observed. Family samples are key in studies of Mendelian disorders but can also have advantages for studies of complex traits [Laird and Lange, 2006, 2008; Ott et al., 2011]. For example, they can be more robust to population stratification (which may be more acute in rare variant association studies [Gravel et al., 2011]), allow checks for genotyping errors (improving data quality [Abecasis et al., 2001,2002]) and can be enriched for variants of large effect by focusing on families with multiple individuals with extreme phenotypes. Early tests for family-based association [Abecasis et al., 2000; Laird et al., 2000; Laird and Lange, 2008] focused on analysis of transmission disequilibrium, but newer tests rely on variance component models [Chen and Abecasis, 2007; Kang et al., 2010] to account for stratification, resulting in tests of association that are typically more powerful [Chen and Abecasis, 2007]. Our work also builds on computational enhancements in methods for variance component analysis, which have now been extended to samples of unrelated individuals (using empirical kinship matrices, estimated from genotype data) [Kang et al., 2010; Lippert et al., 2011; Zhou and Stephens, 2012].

Here, we describe family-based association tests for rare variants that allow analysis of quantitative traits, with or without covariates, and show how these tests can be applied in meta-analysis settings. Our methods are based on the insight that gene-level test statistics can be constructed from single-variant score statistics and estimates of the covariance between those [Liu et al., 2014]. We first analyze single variants using efficient computational algorithms for evaluation of variance component models [Lippert et al., 2011]. We then develop family-based burden (weighted and un-weighted), sequence kernel association tests (SKAT), and variable frequency threshold (VT) tests. Using simulation, we show that type I error is well controlled and compare different testing approaches. As expected, SKAT tests are more powerful when the fraction of associated variants in each gene is small or associated rare variants have opposite directions of effect; VT tests are more robust to the choice of allele frequency threshold for grouping variants. Our analysis of exome chip genotypes and from the HUNT and SardiNIA studies shows that our methods are well calibrated and powerful enough to identify several signals at lipid-associated loci.

There has been much recent work focused on extending gene-level association tests to families. Examples include various family-based burden tests [De et al., 2013; Saad and Wijsman, 2014; Schaid et al., 2013] and variance component based tests [Chen et al., 2013; Ionita-Laza et al., 2013; Saad and Wijsman, 2014; Schaid et al., 2013; Schifano et al., 2012; Svishcheva et al., 2014]. A key difference in our implementation, compared to previous work [Chen et al., 2013; Saad and Wijsman, 2014; Schaid et al., 2013; Schifano et al., 2012; Svishcheva et al., 2014], is that we construct our gene-level statistics using single-variant statistics as input. This allows us to quickly re-evaluate gene-level statistics when gene definitions changes, makes it practical to implement variable frequency threshold based tests, and facilitates meta-analyses. To ensure computational efficiency in genome-wide analyses, our implementation uses a score test that requires fitting a maximum likelihood model only once (rather than a Wald test that would require it for every gene [Saad and Wijsman, 2014]). We also focused on methods that could accommodate a diverse mix of family structures or even samples that include both families and unrelated individuals. This is in contrast to transmission-based tests, such as proposed by [De et al., 2013; Ionita-Laza et al., 2013] that are limited to simpler family structures and cannot account for cryptic relatedness. As usual, we expect transmission-based tests may provide greater protection against stratification—but at the cost of greatly reduced power.

We characterize settings where family studies can provide greater power to detect rare variants with moderate to large phenotypic consequences than studies of unrelated individuals. In studies of unselected samples, this is due to a "Jackpot" effect, whereby a family sample may more easily include multiple copies of trait-associated rare alleles in one locus. While for each locus, the expected number of rare alleles will be the same in a family sample or an unrelated sample of same size, family samples are much more likely to exceed this expectation by a large amount. Our simulations show that this difference can have a large impact on power. All the methods described here are implemented in freely available C++ code and tools.

Methods

In this section, we first describe a variance component model to handle familial relationships. Then, we describe how single-variant association statistics and their covariance matrices can be calculated and how gene-level association tests can be constructed. Next, we describe meta-analytic approaches for both single-variant and gene-level association tests. Finally, we discuss the computational cost of our proposed approach and provide practical suggestions to improve computational performance.

Modeling Relatedness

In a sample of n individuals, we model the observed phenotype vector ($\mathbf{y}$) as a sum of covariate effects (specified by a design matrix $\mathbf{X}$ and a vector of covariate effects $\boldsymbol{\beta}$), additive genetic effects (modeled in vector $\mathbf{g}$ and nonshared environmental effects (modeled in vector ε). Thus,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g} + \varepsilon \qquad (1)$$

We assume that genetic effects are normally distributed, with mean $\mathbf{0}$ and covariance $2\sigma_g^2 \mathbf{K}$, where the matrix $\mathbf{K}$ summarizes kinship coefficients [Lange, 1997] between sampled individuals and σ_g^2 is a positive scalar describing the genetic contribution to the overall variance. We assume that nonshared environmental effects are normally distributed with mean $\mathbf{0}$ and covariance $\mathbf{I}\sigma_e^2$, where $\mathbf{I}$ is the identity matrix.

To estimate **K**, we either use known pedigree structure to define $\hat{\mathbf{K}}$ [Lange, 1997] or else use the Balding-Nicols empirical estimator [Astle and Balding, 2009] that uses observed genotypes to estimate kinship as

$$\hat{\mathbf{K}} = \frac{1}{\nu} \sum_{i=1}^{\nu} \frac{(\mathbf{G_i} - 2f_i 1)(\mathbf{G_i} - 2f_i 1)^{\mathrm{T}}}{4f_i(1-f_i)}$$

where ν is the count of variants, $\mathbf{G_i}$ is a genotype vector where each element encodes the number of observed minor alleles in a particular individual, and f_i is the estimated allele frequency for the ith variant. Model parameters, $\hat{\boldsymbol{\beta}}$, $\widehat{\sigma_g^2}$, and $\widehat{\sigma_e^2}$, are estimated using maximum likelihood and the efficient algorithm described in [Lippert et al., 2011]. For convenience, let the estimated covariance matrix of **y** be $\widehat{\boldsymbol{\Omega}} = 2\widehat{\sigma_g^2}\hat{\mathbf{K}} + \widehat{\sigma_e^2}\mathbf{I}$.

Single-Variant Association Tests and Summary Statistics

Since our gene-level association tests will build on single-variant test statistics [Chen and Abecasis, 2007], we will first describe single-variant test statistics and their corresponding variance-covariance matrix.

Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma_i \left(\mathbf{G_i} - \overline{\mathbf{G_i}}\right) + \mathbf{g} + \varepsilon$$

This model is a refinement of Equation (1) above, adding a scalar parameter γ_i to measure the additive genetic effect of the ith variant. As usual [Lange, 1997], the score statistic for testing $\mathbf{H}_0 : \gamma_i = 0$ is as follows:

$$U_i = \left(\mathbf{G_i} - \overline{\mathbf{G_i}}\right)^{\mathrm{T}} \widehat{\boldsymbol{\Omega}}^{-1} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

And the variance-covariance matrix of these statistics is as follows:

$$\mathbf{V} = \left(\mathbf{G} - \bar{\mathbf{G}}\right)^{\mathrm{T}} \left(\widehat{\boldsymbol{\Omega}}^{-1} - \widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\widehat{\boldsymbol{\Omega}}^{-1}\right)\left(\mathbf{G} - \bar{\mathbf{G}}\right)$$

Under the null, test statistics, $T_i = \frac{U_i^2}{V_{ii}}$, are asymptotically distributed as chi-squared with one degree of freedom.

Gene-Level Association Tests for Family Samples

Using single-variant statistics U_i and their variance-covariance matrix **V**, we are now ready to construct a variety of gene-level association test statistics that combine information across variants.

The simplest statistic for a burden test is to estimate the average genetic effect across a series of variants satisfying certain functional (e.g., nonsynonymous or protein truncating variants) and frequency criteria (e.g., allele frequency <0.05). Then the rare variant burden for each individual can be defined as a weighted sum of allele counts for variants satisfying these criteria. Abstractly, we define the rare variant burden as $(\mathbf{G} - \bar{\mathbf{G}})\mathbf{w}$, where $\mathbf{w} = (w_1, w_2, \ldots, w_m)^T$ is a vector of weights for each of the m variants in the gene. A regression parameter measuring the average effect of each variant can be estimated using the model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma \left(\mathbf{G} - \bar{\mathbf{G}}\right)\mathbf{w} + \mathbf{g} + \varepsilon$$

To test the null hypothesis $\gamma = 0$, we use a score statistic, expressed as a function of single-variant statistics $\mathbf{w}^{\mathrm{T}}\mathbf{U}$ with variance $\mathbf{w}^{\mathrm{T}}\mathbf{V}\mathbf{w}$.

Then the burden test statistic $\mathrm{T}_{\mathrm{burden}} = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{U}}{\sqrt{\mathbf{w}^{\mathrm{T}}\mathbf{V}\mathbf{W}}}$ is asymptotically normal with mean zero and variance one.

Variable Threshold Tests for Family Samples

The simplest burden tests will be effective when appropriate frequency thresholds and functional annotation are used to select functional variants for analysis. However, this is challenging to do because the optimal frequency thresholds will vary by gene and by phenotype [Lange et al., 2014]. One possibility is to define a test statistic that considers many alternative frequency thresholds [Lin and Tang, 2011; Price et al., 2010].

Following the suggestions of Price et al. [2010] and Lin et al. [2011], we will define the variable threshold test statistic as the maximal absolute value of burden test statistics across all possible frequency thresholds, $\mathrm{T}_{\mathrm{VT}} = \max_F |\mathrm{T}_{\mathrm{burden}F}|$, where $\mathrm{T}_{\mathrm{burden}F} = \frac{\phi_{\mathrm{F}}^{\mathrm{T}}\mathbf{U}}{\sqrt{\phi_{\mathrm{F}}^{\mathrm{T}}\mathbf{V}\phi_{\mathrm{F}}}}$ is the burden test statistic calculated with frequency threshold F and ϕ_{F} is a vector of 0s and 1s indicating whether a variant has allele frequency below F. Burden statistics calculated using different frequency thresholds jointly follow a multivariate normal distribution with mean **0**, and variance-covariance matrix $\psi_{\mathrm{ij}} = \frac{\phi_{\mathrm{i}}^{\mathrm{T}}\mathbf{V}\phi_{\mathrm{j}}}{\sqrt{\phi_{\mathrm{i}}^{\mathrm{T}}\mathbf{V}\phi_{\mathrm{i}}}\sqrt{\phi_{\mathrm{j}}^{\mathrm{T}}\mathbf{V}\phi_{\mathrm{j}}}}$ [Lin et al., 2011]. P-values can be evaluated using the cumulative density function of this multivariate normal distribution [Genz, 1992].

Sequence Kernel Association Tests

Another refinement is to use a test statistic that allows for variants in the same gene to modify the phenotype in opposite directions [Chen et al., 2013; Ionita-Laza et al., 2013; Wu et al., 2011; Yan et al., 2014]. For example, in some genes [Abifadel et al., 2003], both gain-of-function and loss-of-function alleles have been described and these signals might cancel each other in a standard burden analysis. The model for this type of test is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma \left(\mathbf{G} - \bar{\mathbf{G}}\right) + \mathbf{g} + \varepsilon$$

In this alternative model, the single-variant effects γ_i are assumed to follow a shared distribution, with mean 0 and variance τw_i. We test the null hypothesis of no association using the statistic $\mathrm{T}_{\mathrm{SKAT}} = \mathbf{U}^{\mathrm{T}}\mathbf{W}\mathbf{U}$ to evaluate whether τ is nonzero [Chen et al., 2013; Wu et al., 2011]. As usual, $\mathbf{W} = \mathrm{diag}(w_1, w_2, \ldots, w_m)$ is a diagonal matrix indicating the weight of each variant. T_{SKAT} is distributed as a mixture chi-squared with weights $\lambda_1, \lambda_2, \ldots, \lambda_n$ corresponding to the eigenvalues of $\mathbf{V}^{\frac{1}{2}}\mathbf{W}\mathbf{V}^{\frac{1}{2}}$ and the χ_1^2 (i) correspond to independently distributed chi-squared variables, each with one degree of freedom [Wu et al., 2011]. P-values can be

evaluated using the Davies algorithm [Davies, 1980] or a moment matching algorithm [Liu et al., 2009].

Meta-Analysis

Since we derived all the statistics above from single-variant score statistics and their covariance matrix, our approach can be readily extended to meta-analyses. We first define the overall single-variant score statistics and their variance-covariance matrix as $U_{\text{meta}_i} = \sum_{k=1}^{s} U_{ik}$ and $V_{\text{meta}_j} = \sum_{k=1}^{s} V_{ij,k}$, where U_{ik} and $V_{ij,k}$ are the single-variant score statistic and variance-covariance matrix components from study k and s are the total number of studies. Whenever variant i is unobserved in study k, we set $U_{ik} = 0$ and $V_{ij,k} = 0$ for all j. Next, we simply calculate burden, VT, and SKAT meta-analysis statistics using the formulae above.

Computational Efficiency

Since we rely on score statistics and their covariance, we only need to fit the linear mixed model once under the null hypothesis. Fitting parameters for this null mixed model is a major part of the computational cost of our approach. Standard EM or Newton-Raphson methods require calculating the inverse of the covariance matrix in each iteration—with time complexity $O(n^3)$, too costly for large datasets. Instead, we used the computationally efficient algorithm described in Lippert et al. [2011] to estimate the variance components and fixed effects under the null (Eq. (1)). The algorithm begins with a one-time singular value decomposition (SVD) of the relationship matrix $\hat{\mathbf{K}}$, a step which has time complexity $O(n^3)$). The results of this decomposition are used in a factorization that transforms the phenotype vector and design matrix so that transformed phenotypes are identically and independently distributed. This second step has time complexity $O(n^2)$. After transformation, the cost of updating the log likelihood becomes linear with respect to sample size n (instead of $O(n^3)$ using the standard approach). Calculating the score statistics and their covariance for all single variants simply requires a transformation of genotypes and has time complexity $O(mn^2)$ for a dataset with m variants. In reality, we calculate covariance of score statistics from markers within a sliding window. For large samples, calculating the SVD of $\hat{\mathbf{K}}$ is the computationally most expensive step. A similar idea with comparable computational efficiency has also been described in Zhou and Stephens [2012]. Both ideas build upon the algorithm described by Kang et al. and implemented in EMMAX [Kang et al., 2010].

When variants are grouped in gene-level tests, the computational cost of calculating the combined test statistics is small after single variants have been analyzed. Obtaining P-values corresponding to these statistics, especially for SKAT and VT analyses, can still be challenging when the number of rare variants in a gene is large. To speed up this step, we used computationally efficient algorithms to evaluate the multi-

variate normal probabilities [Genz, 1992] and the mixture chi-squared distribution [Davies, 1980].

Simulation

We carried out a series of simulations to evaluate the performance of our method. We first simulated a set of 1,000 base-pair sequences, which is close to the length of an average protein coding sequence in humans, using the coalescent (as implemented in the program ms [Hudson, 2002]) and a demographic model calibrated to mimic European population history [Adams and Hudson, 2004; Novembre et al., 2008]. We then carried out gene-dropping simulations [Abecasis et al., 2002] using these simulated sequences as founder haplotypes that were propagated through various pedigree structures (Fig. 1).

To evaluate power, we assigned a fraction of variants below a desired frequency threshold (<0.01 in simulations unless addressed otherwise) as causal. Typically, we assigned minor alleles at causal variants to all have effects in the same direction but, in some cases, a fraction of causal minor alleles was assigned effects in the opposite direction. When assigning effect sizes to causal variants, we considered two trait-generating models—an equal variance model (where the effect size for each variant is proportional to $\frac{1}{\sqrt{p(1-p)}}$, a function of the allele frequency p that ensures each causal variant explains the same amount of trait variance) and an equal effect-size model (where the effect size is the same for all causal variants, irrespective of allele frequency). In the equal effect-size model, relatively common variants explain a larger amount of the variance; while in the equal variance model, rarer variants have larger effect sizes (see supplementary Fig. S1 for demonstration). Genetic effects were set so that the total variance explained by each gene (h^2_{gene}) was in the 0.1–2% range. Empirical power was calculated using 10,000 simulations for each parameter combination. We used $\alpha = 1 \times 10^{-8}$ for single-variant association power and $\alpha = 2.5 \times 10^{-6}$ for gene-level association power. Type I error rate for gene-level tests was estimated using 5,000,000 simulations. To compare studies of families and unrelated individuals, we held the number of genotyped (or sequenced) individuals constant and compared our ability to detect associated variants in studies using different sampling units. In simulations and following association analysis, kinship matrices estimated from pedigree were used to fit the null linear mixed model.

SardiNIA and HUNT Samples Description

To further evaluate how our method performs in real data analysis, we used exome chip data from the HUNT [Holmen et al., 2014a,b] and SardiNIA [Pistis et al., 2014; Pilia et al., 2006] studies, which genotyped 5,803 and 6,602 individuals, respectively. Here, we analyze HDL, adjusted for age and sex (supplementary Table S3). Genotypes were called using the Illumina GenCall algorithm in combination with zCall V2.2. Detailed quality control (QC) procedures can be found in
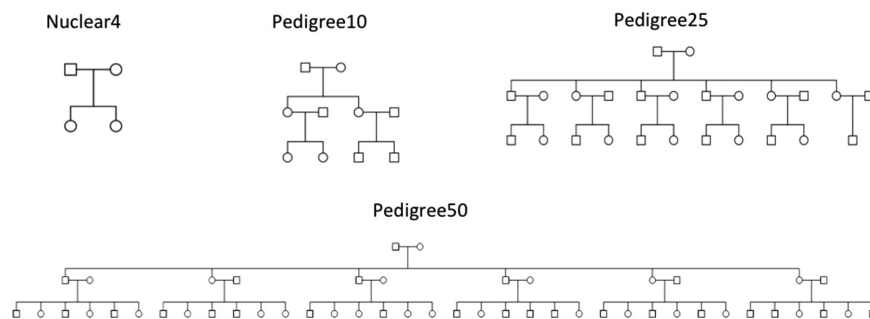
Figure 1. Pedigree structures used in simulations.

[Holmen et al., 2014a] for the HUNT study and [Pistis et al., 2014] for the SardiNIA study.

Results

Type I Error Rate

To evaluate type I error rate, we simulated family samples of 1,000 or 5,000 individuals with family structures matching three-generation pedigrees with 10 (Pedigree10) or 50 (Pedigree50) individuals (see Fig. 1 for details). Within each gene, variants with frequency <0.01 were grouped for analysis. Each type I error estimate summarizes results from 5-million simulations. Supplementary Table S1 shows that the type I error of our gene-level association tests is well controlled for a variety of pedigree structures. Empirical error rates are a little below nominal levels when sample sizes are small ($N = 1,000$), but approach nominal significance as sample size increases ($N = 5,000$).

Power of Different Rare Variant Association Tests

Next, we evaluated the power of our proposed association tests under various scenarios. We used significance level $\alpha = 2.5 \times 10^{-6}$, which corresponds to Bonferroni adjustment for testing of 20,000 genes. We first simulated samples of 5,000 individuals distributed in three-generation pedigrees with 10 individuals each (Pedigree10 in Fig. 1). Variants with minor

allele frequency (MAF) <1% (<5% where noted) explained 1% of the variance in a simulated quantitative trait. When all associated variants had the same effect size and the proportion of causal variants was small (~20%), SKAT had the largest power. When this proportion grew larger (~80%), VT became the most powerful test (Table 1). Although we did not simulate a relationship between frequency and effect size among causal variants, VT provided greater power because it sometimes excluded relatively common unassociated variants from consideration, reducing noise. When fraction of causal variants is small, methods that explicitly allow for heterogeneity in effect sizes do better, since no correlation between causality and effect size was simulated, VT cannot easily exclude most of the unassociated variants. In practice, the true list of causal variants is usually unknown; and allele frequency is often a good proxy to identify variants likely to modify gene function [Nelson et al., 2012]. In a simplified scenario where only causal variants were grouped and other variants were discarded, the basic burden test became optimal (Table 1).

We next considered more complex scenarios. When 20% causal variants decreased trait values and the remainder increased trait values, the power of burden and VT tests dropped dramatically and SKAT became the most powerful test, regardless of the proportion of causal variants (Table 2). When we setup our simulation so that each variant explained the same fraction of trait variance (and, thus, so that rarer variants had larger effects), SKAT remained the most powerful test when the proportion of causal variants was

Table 1. Power when causal variants all increase trait values and have the same effect sizes

| MAF cutoff | Causal percentage (%) | Group by MAF cutoff | | | | Group of only causal variants[b] | | | |
		Burden	Madsen-Browning	VT	SKAT[a]	Burden	Madsen-Browning	VT	SKAT
0.01	20	9.7	3	13.1	36.6	94.3	86.7	92.9	82.6
	80	82.4	64.7	88.1	61	96	82.1	94.3	70.7
0.05	20	14.6	2.6	24.9	36.3	95.4	75.3	93.8	86.5
	80	81.3	39.5	89.2	75	96.3	55.3	94.3	82.9

Simulated samples each had 5,000 individuals, organized in families with Pedigree10 structure (see Fig. 1). Causal variants were selected among those identified in simulated 1,000 base-pair sequences and explained 1% of trait variance. Each causal variant had the same effect size and direction. Power is tabulated as a percentage of simulations exceeding significance threshold. Significance level $\alpha = 2.5 \times 10^{-6}$ was used in all simulations.
[a] Power calculated from Madsen-Browning weighted SKAT.
[b] Power when grouping only causal variants. This column represents the largest power we can achieve for each simulation setting.

Table 2. Power comparison when causal variants can have opposite effects

MAF cutoff	Causal percentage (%)	Group by MAF cutoff				Group of only causal variants			
		Burden	Madsen-Browning	VT	SKAT	Burden	Madsen-Browning	VT	SKAT
0.01	20	4.6	0.4	6.0	36.7	38.9	21.1	43.4	83.2
	80	30.5	10.4	33.4	60.0	42.6	18.8	42.2	69.0
0.05	20	11.7	1.3	15.0	35.7	55.4	22.3	58.3	88.3
	80	44.0	7.8	47.1	74.7	55.1	12.2	54.3	81.6

Simulated samples each had 5,000 individuals, organized in families with Pedigree10 structure (see Fig. 1). Causal variants were selected among those identified in simulated 1,000 base-pair sequences and explained 1% of trait variance. Among causal variants, 20% were randomly selected to be trait-decreasing, and the rest causal variants were trait-increasing. Power is tabulated as a percentage of simulations exceeding significance threshold. Significance level $\alpha = 2.5 \times 10^{-6}$ was used in all simulations.

Table 3. Power comparison when causal variants all increase trait values and explain the same amount of trait variance

MAF cutoff	Causal percentage (%)	Group by MAF cutoff				Group of only causal variants			
		Burden	Madsen-Browning	VT	SKAT	Burden	Madsen-Browning	VT	SKAT
0.01	20	4.3	4.2	9.1	20.8	88.7	94.9	90.8	67.0
	80	66.9	86.6	85.4	20.1	85.5	97.1	93.8	27.0
0.05	20	3.8	5.1	9.3	9.8	78.8	98.0	90.1	53.0
	80	38.6	88.5	82.1	9.4	56.0	97.9	92.6	12.4

Simulated samples each had 5,000 individuals, organized in families with Pedigree10 structure (see Fig. 1). Causal variants were selected among those identified in simulated 1,000 base-pair sequences and explained 1% of trait variance. Each causal variant explained the same amount of trait variance. All causal variants were trait-increasing. Power is tabulated as a percentage of simulations exceeding significance threshold. Significance level $\alpha = 2.5 \times 10^{-6}$ was used in all simulations.

small, but the Madson-Browning (MB) weighted burden test outperformed VT and SKAT when the proportion of causal variants was large (80%; Table 3). This was expected since, in this setting, relative effect sizes match those predicted by the MB-weighting scheme.

Power When Misspecifying Frequency Threshold

We next investigated the impact of misspecifying frequency thresholds during analysis. Supplementary Figure S2A shows that when causal variants have the same effect sizes, VT and MB-weighted burden tests perform well as long as the frequency cutoff used during analysis is larger than the cutoff used for simulation. In contrast, the power of SKAT and simple burden tests is greatly reduced when incorrect frequency thresholds are used for analysis. Supplementary Figure S2B shows that when rare causal variants have larger effects and all variants explain the same amount of trait variance, all tests reach maximum power at a frequency threshold less than or equal to 0.01, the threshold for simulating causal variants. Whereas the power of VT and MB remain close to optimal, the power of SKAT and the simple burden tests drops greatly as the frequency threshold used for analysis increases and noncausal and small effect variants enter the analysis. In real data analysis, because true disease model is unclear, we recommend multiple frequency thresholds should be used when using SKAT or simple burden tests [Lange et al., 2014].

Relative Power of Family Samples and Unrelated Individuals

We used simulations to compare the benefits of samples of families and unrelated individuals in association studies. Family samples can allow many copies of the same trait-associated rare alleles to be observed in a single study. Variability in allele counts is larger in families, particularly in pedigrees with many descendants for each founder. For example, for a variant with allele frequency 0.0005 (approximately five alleles expected when 5,000 individuals are sequenced), the standard deviation of the allele counts in a sample matching Pedigree50 (from Fig. 1) is greater than three times larger than a sample of unrelated individuals (see supplementary Table S2 for details)—meaning that the chance of observing >10 copies of the variant is 20% when families matching Pedigree50 are sampled, but 4% in samples of unrelated individuals.

We speculated that the increased variability in allele counts in family samples would mean that family samples might sometimes hit a "jackpot" and sample many copies of a trait-associated rare allele, increasing power. This speculation was supported by our simulations: a sample of 5,000 individuals in families matching Pedigree50 provides greater than twofold greater power to detect a variant with frequency 0.001 and effect size 1 than a population sample of the same size (power was 0.9% in sample of unrelated individuals and 2.3% in sample of families, supplementary Fig. S3). This increase may seem paltry, but it is important to remember that many susceptibility loci underlie each human complex trait: if there are hundreds of such loci and power increases from 0.9% to 2.3% at each of those, the odds of a successful discovery will increase dramatically. The idea of "jackpot" effect was also supported by close examination of our simulation results. Among all 10,000 simulated samples, the average frequency of trait-associated alleles was 0.0010, but in samples that have association P-value $<1 \times 10^{-8}$, the frequency of trait-associated alleles was higher, averaging 0.0032, a greater than threefold increase. The relative advantages of family
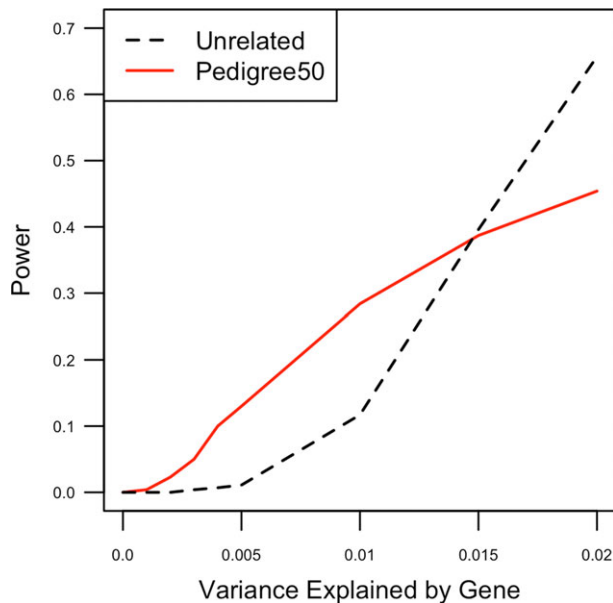
Figure 2. Power to detect gene-level association in family and population samples. All samples had 5,000 individuals. All family samples used the Pedigree50 structure (see Fig. 1 for details). In every simulation, 10,000 haplotypes were simulated and 20% of variants with MAF < 0.01 were randomly selected as causal variants, each explaining the same amount of trait variance. Then, a subset of simulated haplotypes was selected as founder haplotypes, segregated through families according to Mendel's laws, and used to simulate quantitative traits. Power of the SKAT test was evaluated using 10,000 simulations and significance level $\alpha = 2.5 \times 10^{-6}$.

Figure 3. Power to detect gene-level association as a function of pedigree structure. In each simulation, 20% of variants with MAF < 0.01 were randomly assigned as causal, each explaining the same amount of trait variance. Together, causal variants explained 0.5% of trait variance. For comparison, the red line shows the power for one variant with frequency of 0.5 and explaining 0.5% of the trait variance. Power of the SKAT test was evaluated using 10,000 simulations and significance level $\alpha = 2.5 \times 10^{-6}$.

samples over unrelated samples decrease in settings where power (and, typically, the number of expected rare allele carriers) is high. For example, when sample size increases, allele frequency increases, or effect size (or variance explained) increases unrelated samples quickly become more powerful (supplementary Fig. S3).

Consistent to patterns in single-variant association power, Figure 2 shows that family studies have the similar advantages in studies of gene-level rare variant associations. For example, in a sample of 5,000 individuals, power to detect a gene where 20% of variants with frequency <1% are causal and explain 0.5% trait variance increases from 1% for unrelated individuals to 13% for family samples.

Advantages in power from studies of families are strongly correlated to the variance of allele counts (which is a function of family size and pedigree structure). For example, a sample of families matching Pedigree50 (Fig. 1) has largest variance in allele counts (supplementary Table S2) and also the largest power for detecting a gene explaining 0.5% of trait variance in a sample of 5,000 individuals (Fig. 3), whereas a sample of families matching Nuclear4 (Fig. 1) has the smallest variance in allele counts and provides the smallest increase in power relative to samples of unrelated individuals (in this simulation, 20% of variants with frequency 1% were causal). All family samples have larger variance in allele counts than unrelated samples.
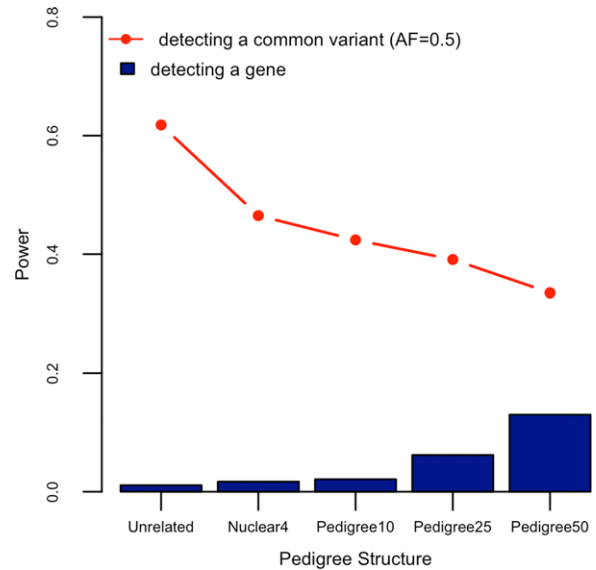
The advantage of family samples extends to extremely rare variants. Figure 4A shows that when 20% of singleton variants (defined as alleles present only once in our initial pool of 10,000 simulated sequences) in a gene were causal explaining 0.5% trait variance, power to detect gene-level association increased dramatically from 3.5% in a study of 5,000 *unrelated* individuals to as much as 19.3% in a study of 5,000 *related* individuals. Figure 4B shows that when sample size increase to 10,000 individuals, the window where family samples are more advantageous becomes narrower.

In all examples highlighted so far, family studies outperform studies of unrelated individuals but in all of these examples power was low for both families and unrelated individuals. We expect that this is actually a common situation in human genetic studies—there may be very large numbers of trait-associated loci but any single study may only provide enough power to detect a few of these. To explore this situation directly, we estimated power to detect at least one of several disease-associated loci. Assuming power to detecting association at a specific gene is p and x genes with similar effect variants exist, then the power to detecting at least one of these is $1 - (1 - p^x)$, assuming independent genes. Figure 5 shows dramatic advantages in the power to detecting at least one of 20 trait-associated genes, each explaining the same proportion of trait variance. For example, power to detecting at least one gene explaining 0.5% trait variance (when 20% variants in the gene and with frequency <1% are causal)
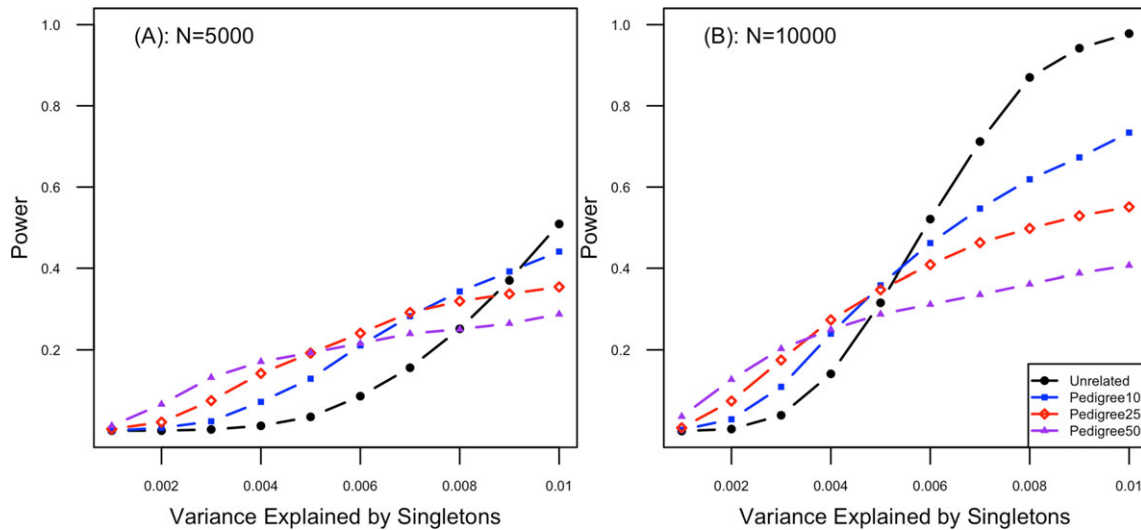
Figure 4. Power to detect gene-level association when singletons are causal. In each simulation, 10,000 simulated haplotypes were simulated. Twenty percent singletons from these haplotypes were chosen as causal variants, together explaining various proportions of trait variance. Trait heritability was 40%. Then, a subset of haplotypes was used to seed founder haplotypes in each family sample. Only singletons or private variants were grouped for association tests. A total of 10,000 simulations were used to evaluate power in samples of 5,000 individuals (panel A) or 10,000 individuals (panel B). See Figure 1 for details of pedigree structures. Power was evaluated in 10,000 simulations using significance level $\alpha = 2.5 \times 10^{-6}$.
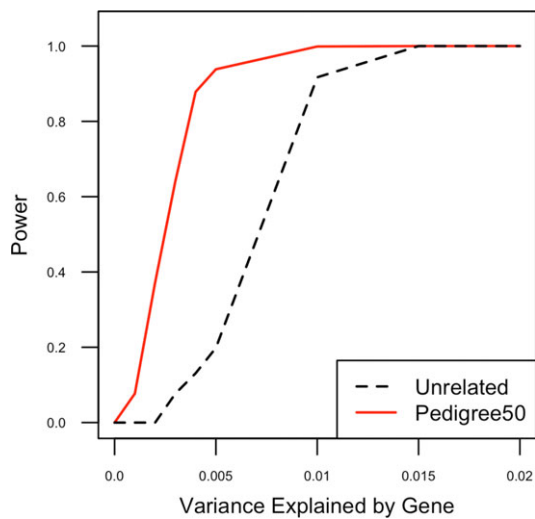

Figure 5. Power to detect at least one of 20 causal genes. Assuming power to detect association at a specific gene is p and n genes with similar effect variants exist, then the power to detect at least one of these is $1 - (1 - p)^n$. See Figure 2 for power to detect a single gene and additional details of simulation settings.
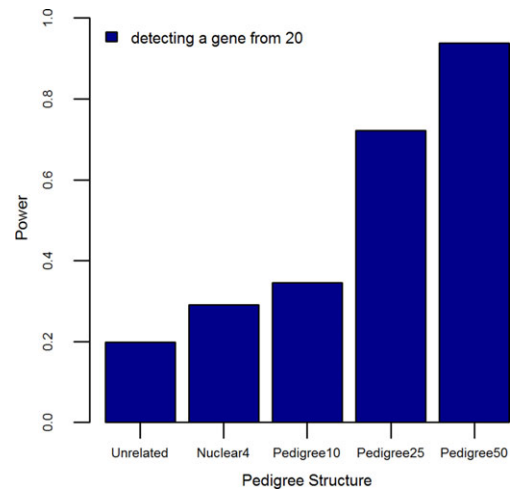

Figure 6. Power to detect at least one of 20 causal genes as a function of pedigree structure. The blue bars show power to detect at least one gene where rare variants explain 20% of trait variance and 20 such genes exist. See the legends of Figure 3 for simulation settings. See the legends of Figure 5 for calculating power to detect at least one of n genes with similar effect variants exist.

when 20 such genes exist is >90% in sample of 5,000 individuals distributed in families matching Pedigree50 (Fig. 1), whereas only ∼20% in a sample of 5,000 unrelated individuals. The power advantage in family samples increases with the variability in allele counts, which in turn is driven by pedigree structure (Fig. 6).

Families matching Pedigree50 are not easy to find. For a more realistic comparison of the power of studies of families and unrelated individuals, we repeated our simulations using the family structures and phenotypes observed in the SardiNIA sample. To preserve the correlation of phenotypes among family members, we started with observed HDL

values together with sex, age, and age-squared as covariates. Supplementary Figure S4 shows that the SardiNIA families provide larger power for discovering rare variants with moderate effect sizes than studies of same numbers of unrelated individuals. For example, the SardiNIA sample provides 1.6% power to detect a variant with frequency 0.0001 and effect of 2.5 trait standard deviation units, whereas unrelated samples provide only 0.05% power (supplementary Fig. S4A). If 100 such variants exist, the SardiNIA sample provides ~80% power to detect at least one, but an equal number of unrelated individuals provides only ~5% power to detect at least one of such a variant (supplementary Fig. S4B). When allele frequency increases (supplementary Fig. S4C–F), the SardiNIA sample is still advantageous when effect sizes are moderate.

Real Data Analysis Using SardiNIA and HUNT Studies

To evaluate our approach further, we meta-analyzed blood HDL levels for 11,556 individuals from the HUNT and SardiNIA studies (see supplementary Table S3 for descriptive statistics for traits). Overall, 93,831 and 76,828 sites were polymorphic in the HUNT and SardiNIA studies, respectively, resulting in 117,958 polymorphic variants when combining the two studies (supplementary Table S4). Among those, 52,700 variants were shared in both studies (supplementary Table S5), 41,130 variants are unique to the HUNT study, and 24,128 variants are unique to the SardiNIA study (supplementary Table S6). Using our meta-analysis method, both shared and nonshared variants contribute to association signals.

We first generated summary statistics for each study adjusting for relatedness using empirical kinship matrices estimated from genotype data. Within each sample, test statistics were well calibrated with genomic control 1.00 in HUNT study and 1.01 in SardiNIA sample (see supplementary Fig. S5 for QQ plots). To illustrate the importance of taking into account phenotype correlations, consider that analyzing the SardiNIA exome chip data and treating the samples as unrelated results in a genomic control value of 1.45, which is unacceptably high (results not shown); but using our approach, genomic control becomes 1.01. We next proceeded to meta-analyze single variants. Supplementary Figure S5 shows that our meta-analysis statistics were also well calibrated with genomic control value <1.05, both for common and rare variants. At a significance threshold of $P < 4.23 \times 10^{-7}$ (corresponding to 0.05/117,958), we found significantly associated low-frequency and rare variants at *CETP, LIPC, LIPG,* and *LPL* for HDL (MAF < 5%; see supplementary Fig. S6 for Manhattan plots). Significant rare variants were only found in *LIPC* and *LIPG* (MAF < 1%).

We then proceeded to gene-level meta-analyses. Again, test statistics appear well calibrated, with genomic control value <1.05 (see supplementary Fig. S7 for QQ plots). Also, by examining QQ plots from SardiNIA and HUNT studies (see supplementary Fig. S7), we discovered that, for family samples or samples from isolated population, in the analysis of rare variants, a small number of individuals can be quite influential such that all variants that are shared between this set of individuals (or families) will exhibit similar and often small *P*-values. This can lead to apparent inflation in QQ plots, where confidence intervals are calculated assuming all statistics are independent. At a significance threshold of $P < 2.84 \times 10^{-6}$ (corresponding to 0.05/17,574 and thus allowing for the number of genes tested), we found association at *APOC3, CETP, LIPC, LIPG,* and *LPL* for HDL (see Table 4 for tabulated results and supplementary Fig. S8 for Manhattan plots). Among those, *APOC3, LIPG,* and *LPL* had evidence of association stronger than the most significant single variant in the region. In *APOC3,* none of the individual low frequency and rare variants had *P*-value lower than 10^{-4} on its own (Table 4).

Comparison with Other Methods and Tools

To validate our approach, we compared our implementation to several others in a simulated family sample of 10,000 individuals distributed across 1,000 families matching Pedigree10 (see Fig. 1). A total of 4,000 genes with 1,000 base-pair were simulated in families from a pool of haplotypes. A quantitative trait was simulated under the null. Variants with MAF < 0.05 were grouped for gene-level tests. Pedigree-based kinship matrices were used in all analyses. We then analyzed the simulated sample using our own famrvtest (SKAT, burden, and VT tests), pedgene (burden and kernel tests) [Schaid et al., 2013], famSKAT [Chen et al., 2013], and FFBSKAT [Svishcheva et al., 2014]. Supplementary Figure S9 shows that all tests generate well-controlled QQ plots under the null.

To compare methods under the alternative, we simulated a dataset of 5,000 individuals (500 × Pedigree10) where a 1,000 base-pair long gene where 50% variants with MAF < 0.05 were causal and together explained 1% trait variance. We simulated datasets where all causal variants had the same direction and also where half of the causal variants had opposite effects. In this simulation, our method always matched or slightly outperformed alternative implementations (see supplementary Fig. S10).

These comparisons also allowed us to evaluate computation performance and requirements for our tool. Wherever possible, we tried to provide faster computation, less memory use, while still allowing for flexible input formats and varied choices of association tests. famrvtest implemented in a C++ command line tool uses computationally efficient algorithms to fit linear mixed models [Lippert et al., 2011], and recognizes pedigree-based kinship estimates as block-diagonal matrices to save computational effort. For our simulated dataset with 10,000 individuals and 164,323 variants distributed across 4,000 genes, analysis with famrvtest required 1.5 hr and 1.3 GB of memory to calculate both SKAT and burden test statistics, a savings of up to 10- to 100-fold relative to alternative tools (see supplementary Table S7).

Table 4. Significant genes[a] from gene-level meta-analysis of HUNT and SardiNIA exome chip data (HDL)

Gene	Burden	Madsen-Browning	VT (actual MAF cutoff)	SKAT[c]	Variants included[d]	MAF	Effect sizes (SD)	Single-variant P-values
APOC3[b]	2.3×10^{-6}	1.9×10^{-6}	6.4×10^{-6} (6.1×10^{-4})	4.5×10^{-5}	11:116701560:G:A	4.8×10^{-4}	0.959	1.4×10^{-3}
					11:116701353:C:T	5.6×10^{-4}	1.009	1.5×10^{-3}
					11:116701354:G:A	6.1×10^{-4}	0.528	5.7×10^{-2}
CETP	6×10^{-20}	2.7×10^{-3}	2.4×10^{-19} (3.2×10^{-2})	1.2×10^{-20}	16:57015091:G:C	3.2×10^{-2}	−0.359	1.3×10^{-20}
					16:57007387:C:T	4.3×10^{-5}	2.241	2.3×10^{-2}
					16:56995935:C:G	4.3×10^{-5}	−1.572	1.1×10^{-1}
					16:57012039:G:A	4.3×10^{-5}	−0.803	4.2×10^{-1}
					16:57009022:G:A	1.7×10^{-4}	0.309	5.3×10^{-1}
					16:57015076:G:A	2.2×10^{-4}	0.144	7.4×10^{-1}
					16:57012094:A:G	4.3×10^{-5}	0.182	8.5×10^{-1}
LIPG[b]	1.3×10^{-10}	6.7×10^{-9}	4.5×10^{-10} (9.4×10^{-3})	1.9×10^{-8}	18:47109955:A:G	9.4×10^{-3}	0.375	4.5×10^{-8}
					18:47113165:C:T	9.1×10^{-4}	0.668	2.3×10^{-3}
					18:47109939:G:A	1.7×10^{-4}	1.012	3.9×10^{-2}
					18:47101838:G:A	4.3×10^{-5}	1.000	3.1×10^{-1}
LPL[b]	3.7×10^{-11}	4.5×10^{-5}	1.2×10^{-10} (2.0×10^{-2})	2×10^{-11}	8:19813529:A:G	2.0×10^{-2}	−0.273	1.3×10^{-8}
					8:19805708:G:A	1.1×10^{-2}	−0.254	7.5×10^{-5}
					8:19816888:C:T	1.1×10^{-3}	0.234	2.3×10^{-1}
					8:19819628:T:G	4.3×10^{-5}	0.193	8.4×10^{-1}
LIPC	1.8×10^{-4}	1.5×10^{-4}	3.2×10^{-5} (6.2×10^{-3})	1.7×10^{-7}	15:58855748:C:T	6.2×10^{-3}	0.539	4.9×10^{-10}
					15:58837989:G:A	7.4×10^{-4}	0.542	2.5×10^{-2}
					15:58833993:G:A	3.1×10^{-2}	0.054	1.7×10^{-1}
					15:58830716:G:A	8.7×10^{-5}	0.123	8.6×10^{-1}
					15:58853079:A:C	5.9×10^{-3}	−0.003	9.8×10^{-1}
					15:58860956:G:A	4.3×10^{-5}	0.025	9.8×10^{-1}

[a] Significance level 2.84×10^{-6} was used for reporting significant genes. Nonsynonymous, splice, and stop variants with MAF < 0.05 were included in analysis.
[b] The gene-level P-value is smaller than the P-value for each of the single variants included in the test.
[c] P-values of SKAT were generated using weights suggested in Wu et al. [2011].
[d] Variants are in the following format: CHR:POS:REF:ALT.

Discussion

Gene-level association tests and meta-analysis are important tools for discovering rare variant associations. We have proposed a series of methods that facilitate these analyses in family samples (or in samples where cryptic relatedness is modeled using variance components). Our C++ tools implement simple burden tests, weighted or un-weighted; and variable threshold tests as well as SKAT tests that outperform other tests when only small fractions of variants in each gene are causal or when variants with opposite effects reside in the same gene.

We compare the relative benefits of family samples and population samples. By simulation, we show that family samples can provide substantially greater power for rare variant association studies because of a "jackpot" effect—the potential for observing many copies of a trait-associated rare variant. This advantage is likely to be extremely important in the first generation of rare variant association studies, each of which is only expected to detect a small fraction of all the true rare variant association signals. An example of successful discovery of such variant is rs72658864/V578A in gene *LDLR*, a rare variant associated to LDL with effect size 23.7 mg/dl [Sanna et al., 2011]. This variant was observed with frequency 0.00035 in the SardiNIA sample, where it was present in multiple families, but has not yet been observed in the 1000 Genomes [Abecasis et al., 2012] or the NHLBI Exome Sequencing Projects [Fu et al., 2013; Tennessen et al., 2012] suggesting that it is rare indeed.

We demonstrate the utility of our methods by analyzing two samples with complex interrelatedness. Meta-analysis of SardiNIA and HUNT resulted in a well-calibrated genomic control value of 1.02 and increased signal at many loci known to be associated with HDL—demonstrating the feasibility of including family samples in rare variant meta-analysis. We expect that meta-analysis will be useful not only for combining data across studies but also to facilitate analysis of large samples genotyped or sequenced across multiple platforms or analyzed using a single platform but in a batched manner.

We foresee several potential areas for refinement of our methods. For example, a limitation for our current approach to meta-analysis is that cross-study relatedness and sample overlap are not modeled. In genome-wide studies, it may be possible to overcome this limitation by using the genome-wide correlation of test statistics between pairs of studies to calculate an adjustment factor that could account for overlap or relatedness between individuals in two studies [Lin and Sullivan, 2009]—as suggested by [Lin et al., 2011] for single marker meta-analyses. Extension of this idea has also been proposed in [Han et al., 2013]. Extending our methods to noncoding variants will also be attractive, particularly since the majority of trait-associated variants found to date are located in noncoding regions. A difficulty will be the development of good grouping strategies for noncoding variants, where interpretation of functional consequence is more challenging. Another challenge we foresee is the extension of our methods to discrete traits. The natural way to do this is to

consider an underlying continuous liability scale and use multivariate integration to fit the model, but there may be more computationally efficient alternatives to be discovered.

In summary, we have proposed a series of gene-level association tests for family samples and methods for calculating these in a meta-analysis of related and/or unrelated samples. We also implemented our methods in freely available and open source C++ tools: http://genome.sph.umich.edu/wiki/FamRvTest and http://genome.sph.umich.edu/wiki/RAREMETAL. We hope these tools and methods will facilitate the next round of gene-mapping studies.

References

Abecasis GR, Cardon LR, Cookson WO. 2000. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66:279–292.

Abecasis GR, Cherny SS, Cardon LR. 2001. The impact of genotyping error on family-based analysis of quantitative traits. *Eur J. Hum Gene* 9:130–134.

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97–101.

Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073.

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1092 human genomes. *Nature* 491:56–65.

Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, Cruaud C, Benjannet S, Wickham L, Erlich D and others. 2003. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 34:154–156.

Adams AM, Hudson RR. 2004. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. *Genetics* 168:1699–1712.

Astle W, Balding DJ. 2009. Population structure and cryptic relatedness in genetic association studies. *Stat Sci* 24:451–471.

Chen H, Meigs JB, Dupuis J. 2013. Sequence kernel association test for quantitative traits in family samples. *Genet Epidemiol.* 37:196–204.

Chen WM, Abecasis GR. 2007. Family-based association tests for genomewide association scans. *Am J Hum Genet* 81:913–926.

Davies R. 1980. The distribution of a linear combination of chi-square random variables. *J R Stat Soc Ser C Appl Stat* 29:323–333.

De G, Yip WK, Ionita-Laza I, Laird N. 2013. Rare variant analysis for family-based design. *PloS One* 8:e48495.

Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J and others. 2013. Analysis of 6515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493:216–220.

Genz A. 1992. Numerical computation of multivariate normal probabilities. *J Comput Graph Stat* 1:141–150.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Bustamante CD. 2011. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci USA* 108:11983–11988.

Han B. 2013. A general framework for meta-analyzing dependent studies with overlapping subjects in association mapping. http://arxiv.org/abs/1304.8045

Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, Guo Y, Zhang J, Langhammer A, Lochen ML and others. 2014a. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet* 46:345–351.

Holmen TL, Bratberg G, Krokstad S, Langhammer A, Hveem K, Midthjell K, Heggland J, Holmen J. 2014b. Cohort profile of the Young-HUNT Study, Norway: a population-based study of adolescents. *Int J Epidemiol* 43:536–544.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. 2013. Family-based association tests for sequence data, and comparisons with population-based association tests. *Eur J Hum Genet* 10:1158–1162.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42:348–354.

Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, Gupta N, Sklar P, Sullivan PF, Moran JL, et al. 2012. Exome sequencing and the genetic basis of complex traits. *Nature genetics* 44:623–630.

Laird NM, Horvath S, Xu X. 2000. Implementing a unified approach to family-based tests of association. *Genet Epidemiol* 19(Suppl 1):S36–S42.

Laird NM, Lange C. 2006. Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet* 7:385–394.

Laird NM, Lange C. 2008. Family-based methods for linkage and association analysis. *Adv Genet* 60:219–252.

Lange K. 1997. *Mathematical and statistical methods for genetic analysis.* New York: Springer-Verlag.

Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang ZZ, Bizon C, Lange EM, Smith JD, Turner EH and others. 2014. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet* 94:233–245.

Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Christiani DC, Wurfel MM, Lin X. 2012a. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 91:224–237.

Lee S, Wu MC, Lin X. 2012b. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13:762–775.

Lee S, Teslovich TM, Boehnke M, Lin X. 2013. General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet* 93:42–53.

Li B, Leal SM. 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American journal of human genetics* 83:311–321.

Lin DY, Sullivan PF. 2009. Meta-analysis of genome-wide association studies with overlapping subjects. *Am J Hum Genet* 85:862–872.

Lin DY, Tang ZZ. 2011. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* 89:354–367.

Lin DY, Zeng D. 2010. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* 97:321–332.

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. 2011. FaST linear mixed models for genome-wide association studies. *Nat Methods* 8:833–835.

Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, Nikpay M, Auer PL, Goel A, Zhang H and others. 2014. Meta-analysis of gene-level tests for rare variant association. *Nat Genet* 46:200–204.

Liu H, Tang Y, Zhang H. 2009. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Comput Stat Data Anal* 53:853–856.

Madsen BE, Browning SR. 2009. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 5:e1000384.

Marth GT, Yu F, Indap AR, Garimella K, Gravel S, Leong WF, Tyler-Smith C, Bainbridge M, Blackwell T, Zheng-Bradley X, et al. 2011. The functional spectrum of low-frequency coding variation. *Genome biology* 12:R84.

Morris AP, Zeggini E. 2010. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol* 34:188–193.

Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu SA, Fraser D and others. 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14002 people. *Science* 337:100–104.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR and others. 2008. Genes mirror geography within Europe. *Nature* 456:98–101.

Ott J, Kamatani Y, Lathrop M. 2011. Family-based designs for genome-wide association studies. *Nat Rev Genet* 12:465–474.

Pistis G, Eleonora P, Vrieze SI, Sidore C, Steri M, Danjou F, Busonero F, Mulas A, Zoledziewska M, Maschio A and others. 2014. Rare variants genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. *Eur J Hum Genet* doi:10.1038/ejhg.2014.216.

Pilia G, Chen WM, Scuteri A, Orru M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P and others. 2006. Heritability of cardiovascular and personality traits in 6148 Sardinians. *PLoS Genet* 2:e132.

Price AL, Kryukov GV, deBakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. 2010. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* 86:832–838.

Saad M., Wijsman EM. 2014. Power of family-based association designs to detect rare variants in large pedigrees using imputed genotypes. *Genet Epidemiol* 38:1–9.

Sanna S, Li B, Mulas A, Sidore C, Kang HM, Jackson AU, Piras MG, Usala G, Maninchedda G, Sassu A and others. 2011. Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet* 7:e1002198.

Schaid DJ, McDonnell SK, Sinnwell JP, Thibodeau SN. 2013. Multiple genetic variant association testing by collapsing and kernel methods with pedigree or population structured data. *Genet Epidemiol* 37:409–418.

Schifano ED, Epstein MP, Bielak LF, Jhun MA, Kardia SL, Peyser PA, Lin X. 2012. SNP set association analysis for familial data. *Genet Epidemiol* 36:797–810.

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU and others. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345.

Svishcheva GR, Belonogova NM, Axenovich TI. 2014. FFBSKAT: fast family-based sequence kernel association test. *PloS One* 9:e99407.

Tang ZZ, Lin DY. 2013. MASS: meta-analysis of score statistics for sequencing studies. *Bioinformatics* 29:1803–1805.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G and others. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337: 64–69.

Willer CJ, Li Y, Abecasis GR. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190–2191.

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 89:82–93.

Yan Q, Tiwari HK, Yi N, Lin WY, Gao G, Lou XY, Cui X, Liu N. 2014. Kernel-machine testing coupled with a rank-truncation method for genetic pathway analysis. *Genet Epidemiol* 38:447–456.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821–824.