8-2014

General Approaches for Combining Multiple Rare Variant Associate Tests Provide Improved Power Across a Wider Range of Genetic Architecture

Nathan L. Tintle
Dordt College, nathan.tintle@dordt.edu

Brian Greco
University of Michigan

Allison Hainline
Vanderbilt University

Keli Liu
Stanford University

Jaron Arbet
Winona State University

See next page for additional authors

General Approaches for Combining Multiple Rare Variant Associate Tests Provide Improved Power Across a Wider Range of Genetic Architecture

Abstract
In the wake of the widespread availability of genome sequencing data made possible by way of nextgeneration technologies, a flood of gene-based rare variant tests have been proposed. Most methods claim superior power against particular genetic architectures. However, an important practical issue remains for the applied researcher—namely, which test should be used for a particular association study which may consider multiple genes and/or multiple phenotypes. Recently, tests have been proposed which combine individual tests to minimize power loss while improving the robustness to a wide range of genetic architectures. In our analysis, we propose an expansion of these approaches, by providing a general method that works for combining an arbitrarily large number of any gene-based rare variant test—a flexibility typically not available in other combined testing methods. We provide a theoretical framework for evaluating our combined test to provide direct insights into the relationship between test-test correlation, test power and the combined test power relative to individual testing approaches and other combined testing approaches. We demonstrate that our flexible combined testing method can provide improved power and robustness against a wide range of genetic architectures. We further demonstrate the performance of our combined test on simulated genotypes, as well as on a dataset of real genotypes with simulated phenotypes. We support the increased use of flexible combined tests in practice to maximize robustness of rare-variant testing strategies against a wide-range of genetic architectures.

Keywords
genetic analysis, genomic variation, case-control studies

Disciplines
Genetics and Genomics | Medicine and Health Sciences | Statistics and Probability

Comments
Poster from presentation given at the 23rd Annual Meeting of the International Genetic Epidemiology Society, Vienna, Austria, August 28-30, 2014.

Abstract for conference (p. 69, http://www.geneticepi.org/meeting-abstracts/) lists title as "A general approach for combining diverse rare variant association tests provides improved power across a wider range of genetic architecture."

Authors
Nathan L. Tintle, Brian Greco, Allison Hainline, Keli Liu, Jaron Arbet, Alejandra Benitez, and Kelsey Grinde

General Approaches for combining Multiple Rare Variant Associate Tests Provide Improved Power Across a Wider Range of Genetic Architecture

Nathan Tintle[1], Brian Greco[2], Allison Hainline[3], Keli Liu[4], Jaron Arbet[5], Alejandra Benitez[6], Kelsey Grinde[7]

1. Dordt College Department of Mathematics, Statistics and Computer Science 2. University of Michigan, Department of Biostatistics 3. Vanderbilt University, Department of Biostatistics, 4. Stanford University, Department of Statistics, 5. Winona State University Department of Mathematics and Statistics 6. Brown University, Department of Applied Mathematics 7. St. Olaf College Department of Mathematics, Statistics, and Computer Science

I. Abstract

Over the past five years, numerous gene-based rare variant tests of association have been proposed, each of which attempt to combine variants within a gene or region of interest into a single association statistic, with a goal of providing more power than a strategy which analyzes each variant separately. Simulation results have shown that many of these individual tests provide good power for particular genetic architectures, but not others. We have developed a general strategy for combining any two or more gene-based rare variant tests using an adaptive approach, which yields a single p-value representing the cumulative evidence for association across the set of gene-based tests. For example this strategy can take any threshold based test and turn it into a variable-threshold test, combine similar tests (similar statistic with alternative weighting strategies), or combine substantially different tests (e.g., burden tests and variance components tests). Using simulation we provide guidance on the tradeoff between power gains and test robustness versus the number of tests being combined, a result which is based on the correlation structure of the tests are under the null hypothesis of no association. Finally, we demonstrate how recent results from our group which suggested a substantially different gene-based test which is robust to high proportions of non-causal variants, combined with other popular tests (burden and variance component tests), can provide improved power across a wider range of genetic architecture.

II. Introduction

- Over the years, many tests of genotype-phenotype association have been proposed , and recent work has shown that different types of these tests are more or less powerful under different genetic architectures.
- Recent work by our group has classified these tests into two groups: Length tests (also known as burden, collapsing, and/or linear tests) and Joint tests (alternatively, variance components or quadratic tests). (Liu et al., 2013)
- Length tests can be powerful when the proportion of causal variants in the region is large and the effects of the causal variants tend to be similar. Joint tests can be more powerful than length tests when there are larger proportions of non-causal variants and there is more variation in the effects of causal variants (e.g., both risk increasing and risk-reducing variants).
- Four recent papers have proposed combining test statistics across both the length and joint classes to yield more powerful test statistics (Derkach et al., 2013; Lee, Wu, et al., 2012; Lee, Emond, et al., 2012; Liu et al., 2013).
- Liu et al. (2013) also showed that length and joint tests can be further classified by the norm, *p*, used in the formulation of the test statistic. To date, most length tests use *p*=1 and most joint tests use *p*=2. Liu et al. demonstrate that higher choices of norm provide increased robustness to large proportions of non-causal variants.
- More general test-combining strategies (such as combinations that include length and joint tests with higher norms) may yield more powerful results when the component tests being combined are powerful for a wide range of genetic architectures.

III. Methods

General strategy for combining tests

- First: compute the *p*-value for each of the *k* different gene-based rare variant tests, yielding a vector of *p*-values, ***p***= (see bottom p.3 of draft) for each gene of interest.
- Use this to generate a test statistic which summarizes the strength of evidence across ***p.*** We use two different test statistics:
  - Fisher's combined *p*-value test statistic: F_k= $F_k = \sum_{i=1}^{k} -2\log(p_i)$
  - Minimum *p*-value: *Min(**p**)*=argmin(***p***).

Description of the permutation strategy

- Derkach et al. (2012) propose an efficient permutation strategy for assessing the significance of *S* which we extend and apply here
- Utilize any number of rare variant tests.
- Find the minimum p-value of the tests..
- Through permutation, empirically find the distribution of minimum p- values.
- Adjust the actual minimum p-value according to the empirical distribution.

Rare variant tests

- List of individual tests considered: *Sequence Kernel Adaptive Test* (*SKAT*); *Sequence Kernel Adaptive Test-Optimal* (*SKAT-O*); *Combined Multivariate and Collapsing Test* (*CMC); Length tests with different norms (L(p)); Joint tests with different norms (J(p)); Odds Ratio Weighted Sum Statistic (ORWSS)*
- List of the combinations explored:

Table 1 – List of combined Tests Used

Length tests with different norms	L(1), L(2), L(4), L(∞)
Joint tests with different norms	J(1), J(2), J(4), J(∞)
Variable threshold ORWSS	ORWSS(>0), ORWSS(<0), ORWSS(≠0), ORWSS(>0), ORWSS(<0), ORWSS(≠0), ORWSS(>0), ORWSS(<0), ORWSS(≠0)
Similar length tests	CMC, L(1)
Similar joint tests	SKAT, J(2)
Typical length-joint combined test	SKAT, CMC
Length and joint tests across norms	L(1), L(2), L(4), L(∞),(J(1), J(2), J(4), J(∞))
Length and joint with some norms	L(1), L(4), J(1), J(4)
Generic length-joint combined test	L(1), J(2)
Heterogeneous combined test #1	CUMIT, ORWSS(>0), ORWSS(<0), ORWSS(≠0), ORWSS(>0), ORWSS(<0), ORWSS(≠0), ORWSS(>0), ORWSS(<0), ORWSS(≠0), L(1), L(2), L(4), L(∞),J(1), J(2), J(4), J(∞)
Heterogeneous combined test #2	SKAT-O, J(∞)
Heterogeneous combined test #3	ORWSS(>0), ORWSS(<0), ORWSS(≠0), ORWSS(>0), ORWSS(<0), ORWSS(≠0), ORWSS(>0), ORWSS(<0), ORWSS(≠0), L(1), L(2), L(4), L(∞),J(1), J(2), J(4), J(∞)

Table 2 – Most Powerful Tests

row.names	ovr_pctbest	ovr_within5
Hetero.2 FISH	0.132	0.817
All Joint FISH	0.183	0.736
Joint2	0.162	0.726
(SKAT,J2) MINP	0.157	0.726
(SKAT,J2) FISH	0.162	0.726
SKAT1in	0.127	0.721
Hetero.3 FISH	0.152	0.706
All Joint MINP	0.107	0.701
(L1,J2) FISH	0.289	0.690
Hetero.2 MINP	0.132	0.675
(SKAT,CMC) FISH	0.152	0.660
rSKAT.stat	0.178	0.645
SKATOlin	0.142	0.640
(L1,J2) MINP	0.137	0.635
Joint4	0.102	0.619
Hetero.3 MINP	0.127	0.619
(SKAT,CMC) MINP	0.137	0.589
Joint1	0.122	0.563
All Length and Joint FISH	0.127	0.543
(L1,L4,J1,J4) MINP	0.132	0.543
All Length and Joint MINP	0.132	0.528
(L1,L4,J1,J4) FISH	0.147	0.523
Hetero.1 FISH	0.173	0.492
Hetero.4 FISH	0.137	0.482
ourORWSS0	0.081	0.421
JointInf	0.010	0.401
ourORWSS0.5	0.010	0.376
Hetero.1 MINP	0.122	0.350
Hetero.4 MINP	0.117	0.345

Table 1. Left column gives name of combined test, right column lists the individual tests that make up the combined test

Table 2. Lists percent of 197 simulations where a particular combined or individual test had the highest power, or was within 5% of the most powerful test for that simulation setting. The tests were ranked by how often they were within 5% of the most powerful test.

A) Simulations

Simulation Study # 1: *Investigating the behavior of Min(p) and Fisher's*

Simulation Study # 2: *Investigating the behavior of combinations of gene-based rare variant tests across different genetic disease models*

- 197 simulation settings, representing all possible combinations of the following parameters:
  (1) Number of single nucleotide variants (SNVs) (32 or 64)
  (2) Proportion of non-causal SNVs (0, ¼, ½, ¾, 7/8, 15/16, 31/32, 63/64, 1)
  (3) Proportion of causal SNVs that increase disease risk (0, ¼, ½, ¾, 1), with the remaining causal SNVs causing a decline in disease risk
  (4) Relative risk of causal, risk increasing SNVs (1.1, 1.5 and 2.0); fixed the relative risk of risk-reducing SNVs at 0.5
  (5) Minor allele frequencies simulated in 3:1 ratio of less common (0.1% population MAF) to more common (1% MAF), spread evenly across all non-causal and causal SNVs.
- 500 samples generated at each simulation setting,
- Each individual and combined test applied to each sample (with separate p-values for Min(p), Fisher's, and Bonferonni for each combined test).
- Empirical power estimates computed as percentage of p-values less than 0.05 across the 500 samples.

IV. Results

Min(p) vs Bonferroni vs Fishers

- Across the 197 simulation settings and 12 combined tests (2364 possibilities), there were only 10 times where power of the Bonferroni approach exceeded the power of the Min(p) approach and power gains were minimal (ranging from 0.002 to 0.004) in these cases.
- Across 197 simulation settings, 36.5% of time Min(p) is more powerful than Fishers
- Minimum p-value increases power when tests are very different, but powerful in certain situations (if only 1 test has a p-value of .001, but the others have p-values that are very high, min-p works well)
- Fisher's increases power when tests of different classes all return low p-values (because it combines the information).

Correlation structure between tests

- The performance of Min(p) and Fishers methods will be affected by the correlation structure between the tests (result of Sim. 1).

Figure 1 – Heat Map of Combined Tests

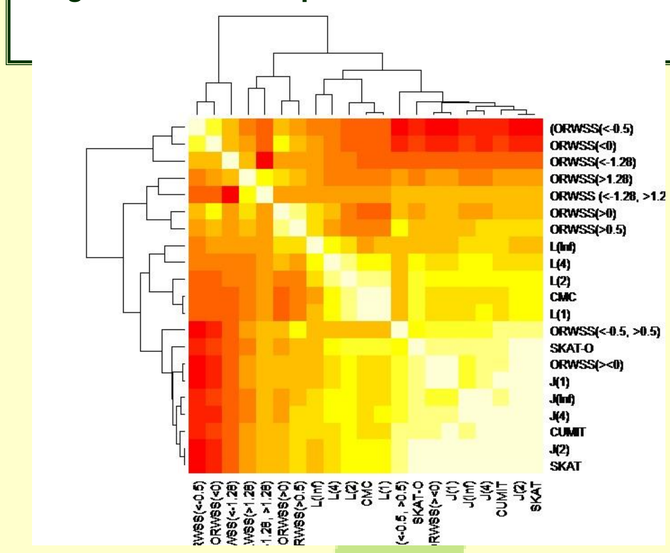

Figure 1. To understand how the p-values between different tests are correlated with each other we created a Heat Map of the p-values between different tests across all simulation settings

Figure 2 – When Fishers generic combo(L1,J2) is more powerful than L1(82.2% of sims) and J2(37.5% of sims)

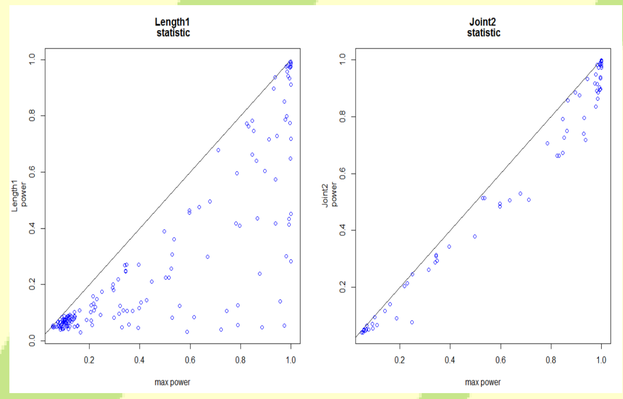

Figure 3 – Hetero2 vs (L1,J2) Power on 80-98.4% Non Causal Simulation Settings

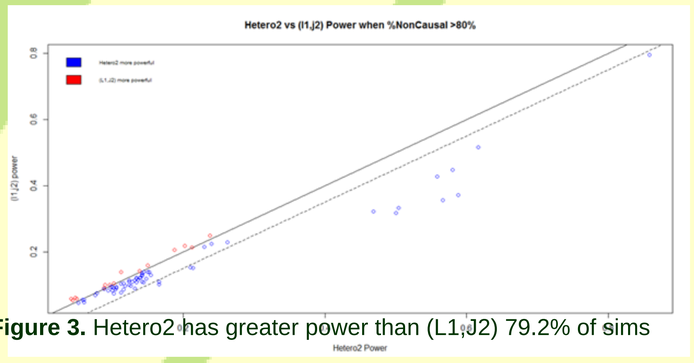

Figure 3. Hetero2 has greater power than (L1,J2) 79.2% of sims

References

Liu, Keli, Shannon Fast, Matthew Zawistowski, and Nathan Tintle. "A Geometric Framework for Evaluating Rare Variant Tests of Association." *Genetic Epidemiology* 37.4 (2013): 345-57. Web. 16 Oct. 2013.

Derkach, Andriy, Jerry F. Lawless, and Lei Sun. "Robust and Powerful Tests for Rare Variants Using Fisher's Method to Combine Evidence of Association From Two or More Complementary Tests." *Genetic Epidemiology* 37.1 (2013): 110-21. Web.