Published in final edited form as:

Genet Epidemiol. 2020 January ; 44(1): 104–116. doi:10.1002/gepi.22273.

An Adaptive Test for Meta-Analysis of Rare Variant Association Studies

Tianzhong Yang^{1, #}, Junghi Kim^{1, #}, Chong Wu², Yiding Ma^{3, 4}, Peng Wei⁴, Wei Pan^{1, *}

¹Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA

²Department of Statistics, Florida State University, Tallahassee, FL, USA

³Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

⁴Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

Single genome-wide studies may be under-powered to detect trait-associated rare variants with moderate or weak effect sizes. As a viable alternative, meta-analysis is widely used to increase power by combining different studies. The power of meta-analysis critically depends on the underlying association patterns and heterogeneity levels, which are unknown and vary from locus to locus. However, existing methods mainly focus on one or only few combinations of the association pattern and heterogeneity level, thus may lose power in many situations. To address this issue, we propose a general and unified framework by combining a class of tests including and beyond some existing ones, leading to high power across a wide range of scenarios. We demonstrate that the proposed test is more powerful than some existing methods in simulation studies, then show their performance with the NHLBI Exome-Sequencing Project (ESP) data. One gene (*B4GALNT2*) was found by our proposed test, but not by others, to be statistically significantly associated with plasma triglyceride. The signal was driven by African-ancestry subjects, but it was previously reported to be associated with coronary artery disease among European-ancestry subjects. We implemented our method in an R package aSPUmeta, publicly available at <https://github.com/ytzhong/metaRV>.

Keywords

aSPU; genetic heterogeneity; whole exome sequencing; whole genome sequencing; statistical power

*Correspondence should be addressed to: Dr. Wei Pan, A460 Mayo Building, MMC 303, Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA; panxx014@umn.edu.

#These authors contributed equally to this work

1 Introduction

As the next-generation sequencing (NGS) and exome chip data become increasingly available in large genetic epidemiology studies, more researchers have shifted their attention from common variants (CVs) with minor allele frequency (MAF) larger than 5% in genome-wide association study to less frequent or rare variants (RVs) with MAF less than 5% or 1% (Cohen et al., 2006; Gibson, 2012; Liu and Leal, 2012a,b; Yano et al., 2016; Liu et al., 2017; Lu et al., 2017; Morrison et al., 2017). RV analysis is complementary to CV analysis in identifying novel genes and locating causal variants (MacArthur et al., 2014; Auer et al., 2014). However, lack of statistical power is a more severe limitation for RV analysis than CV analysis. It requires a large sample size not only to detect the presence of a RV but also to test its possible association with a trait, which, if present, is often weak or modest (Auer and Lettre, 2015). Until recently, there are limited major findings through RV analysis. To further increase power, meta-analysis can be employed by combining information across multiple independent studies.

One of the most important factors that impact the statistical power of meta-analysis is the between-study heterogeneity, i.e., the association strengths may vary across the cohorts. Such heterogeneity is well-acknowledged in the recent development of meta-analysis of the genome-wide association studies (GWAS) (Han and Eskin, 2011, 2012; He et al., 2016) and sequencing studies (Lee et al., 2013; Tang and Lin, 2014, 2015; Fan et al., 2015, 2016), where random effect (RE) models or/and fixed effect (FE) models were considered. In general, RE models relax the assumption of FE models by allowing effects to vary across studies, and they are sometimes more powerful in the presence of moderate to high between-study heterogeneity than FE models. On the other hand, FE models, which assume that genetic effects are the same across all participating studies, are usually more powerful when this assumption is largely met (Han and Eskin, 2011; Lee et al., 2013; Tang and Lin, 2014). Between-study heterogeneity may happen due to a variety of reasons. For example, many RVs have occurred in recent human history and therefore they may not be shared among different populations; a genetic variant can interact with other functional variants or with environmental factors that vary among populations; there may be differences in the study design and implementation, e.g., using different sequencing/genotyping platforms and quality-control criteria. In spite of the well-known between-study heterogeneity, in practice, it is unclear how to best account for it in analyses. Since the between-study heterogeneity may vary from locus to locus, as to be demonstrated in our real data example, it is not straightforward to determine the appropriate type of model to use in advance (Chen and Parmigiani, 2007). Therefore, a new statistical method is needed to accommodate different levels of between-study heterogeneity in practice.

For RV analysis, set-based tests are commonly used, upon which the existing methods proposed for meta-analysis are all built (Lumley et al., 2012; Lee et al., 2013; Tang and Lin, 2014; Liu et al., 2014; Tang and Lin, 2015). These methods combine studies through summary statistics, usually the score statistics and their covariance matrices, but the types of existing set-based tests are limited. One fundamental feature of the set-based tests is that statistical power depends on the true and unknown association pattern. For example, burden tests are more powerful when the genetic effects of the individual RVs are all the same and

in the same direction, whereas variance-component (VC) tests are more powerful when the effects are in different directions (Lee et al., 2014). According to previous studies (Pan et al., 2014), a VC or a burden test may lose power quickly when the association pattern across the RVs becomes sparse; that is, a high proportion of the RVs are neutral and not associated with the trait. As the sequencing depth or sample size keeps increasing, more neutral RVs are expected to be captured even in a causal gene or locus, thus a more comprehensive evaluation of meta-analysis built on different types of set-based tests is critical.

As aforementioned, the association strengths and underlying association patterns across studies/populations influence the power of meta-analysis of RVs, in general, they are unknown while varying across RV sets. Hence, it is compelling to have a highly adaptive test to maintain high power across various scenarios in practice, especially for the purpose of genome-scan. To accommodate different levels of between-study heterogeneity and a variety of association patterns, we took an approach similar to a doubly-adaptive test for pathway analysis (Pan et al., 2015), in which two tuning parameters are used to control the weights on study cohorts and genetic variants. After evaluating a class of tests with different combinations of the tuning parameters, we show that this test is able to approach the most powerful test and thus maintain high power across various scenarios. Later in the paper, we describe our method and its connection with some existing set-based meta-analysis methods. Then, we present a simulation study to examine the operating characteristics of the proposed method in comparison with existing tests, and apply them to the NHLBI Exome Sequencing Project (ESP) and UK10K Project data. In our real-data example, we conduct a trans-ethnic meta-analysis using the ESP data, discovering three novel genes associated with plasma triglyceride levels. The signals in the three loci are driven by African-ancestry (AA) subjects, one of which is identified by our proposed method only. We also demonstrate evidence of increased power by a combined analysis of the UK10K whole-genome sequencing (WGS) data and the ESP exome-sequencing (WES) data. We conclude with a short discussion and provide a publicly available R package implementing our proposed method.

2 Methods

Suppose there are $k = 1, 2, \dots, K$ independent studies. For the k^{th} study, n_k subjects are sequenced in a region with p_k variants. Let X_{kij} denote the j^{th} genotype that takes value of 0, 1, and 2 for i^{th} individual in k^{th} study. We write $X_{ki} = (X_{ki1}, \dots, X_{kip_k})'$ as a vector of p_k genotypes. Let $Z_{ki} = (Z_{ki1}, \dots, Z_{kiqu_k})'$ be a vector of q_k covariates including an intercept, and Y_{ki} denote the trait of interest, which can be continuous or discrete. Through a generalized linear model, Y_{ki} given X_{ki} and Z_{ki} has a conditional mean,

$$g(E(Y_{ki})) = \beta_k' X_{ki} + \alpha_k' Z_{ki}$$

where $g(\cdot)$ is a canonical link function, $\beta_k = (\beta_{k,1}, \dots, \beta_{k,p_k})'$ and $\alpha_k = (\alpha_{k,1}, \dots, \alpha_{k,q_k})'$ are regression coefficients.

Denote the score statistics for testing a study-level null hypothesis $H_{0k}: \beta_k = 0'$ as $U_k = (U_{k,1}, \dots, U_{k,p_k})'$. Each element $U_{k,j}$ measures the association strength between variant j and

the trait in the k^{th} study. Under H_{0k} , U_k is asymptotically normal with mean 0 and covariance matrix V_k . Specifically, for a continuous trait with an identity link, U_k and V_k are obtained as

$$U_k = \frac{1}{\hat{\sigma}_k^2} \sum_{i=1}^{n_k} (Y_{ki} - \hat{\alpha}_k' Z_{ki}) X_{ki}, \quad (1)$$

$$V_k = \frac{1}{\hat{\sigma}_k^2} \left\{ \sum_{i=1}^{n_k} X_{ki} X_{ki}' - \left(\sum_{i=1}^{n_k} X_{ki} Z_{ki}' \right) \left(\sum_{i=1}^{n_k} Z_{ki} Z_{ki}' \right)^{-1} \left(\sum_{i=1}^{n_k} Z_{ki} X_{ki}' \right) \right\}, \quad (2)$$

where $\hat{\sigma}_k^2 = n_k^{-1} \sum_{i=1}^{n_k} (Y_{ki} - \hat{\alpha}_k' Z_{ki})^2$, and $\hat{\alpha}_k = (\sum_{i=1}^{n_k} Z_{ki} Z_{ki}')^{-1} \sum_{i=1}^{n_k} Y_{ki} Z_{ki}'$.

2.1 aSPU-meta test

In meta-analysis, the null hypothesis H_0 is $\beta_1 = \beta_2 = \dots = \beta_K = 0'$ for K study. Tang and Lin (2015) has provided a thorough review of the existing methods for meta-analysis of RVs. The existing methods are usually powerful under a certain combination of genetic architecture and between-study heterogeneity level. A more adaptive test is needed to maintain the statistical power under a wider range of scenarios and carries on the advantage of the existing methods. Without loss of generality, we assume the same set of genetic variants across different studies, i.e., $p_1 = p_2 = \dots = p_K = p$, which is a common practice in meta-analyses for sequencing study. Similar to other methods in the area, we treat the variants that are present in some studies but not in others as of zero effect in the latter studies.

Our proposed meta-analysis method, named aSPU-meta test, is built upon the aSPU test (Pan et al., 2014) that was previously extended to pathway analysis (Pan et al., 2015). Briefly, the aSPU-meta test evaluates a set of SPU tests, each of which has an edge under a certain combination of generic architecture and between-study heterogeneity level, then the aSPU-meta test chooses the most powerful SPU test adaptively. The aSPU-meta test is constructed by the following steps.

1. Obtain the summary statistics U_k and V_k from participating studies. If individual data are available, we can calculate the score statistics U_k and V_k for each study using equations (1) and (2).
2. Construct the gene-based test statistic for each study,

$$S(k; \gamma_1) = \sum_{j=1}^p U_{k,j}^{\gamma_1},$$

where $\gamma_1 \in \Gamma_1$ is a positive integer. $S(k; \gamma_1)$ quantifies the association between a single trait and the set of RVs for k^{th} study.

3. Build a class of SPU(γ_1, γ_2) by combining across all the studies.

$$\text{SPU}(\gamma_1, \gamma_2) = \sum_{k=1}^K [S(k; \gamma_1)]^{\gamma_2},$$

where $\gamma_2 \in \Gamma_2$ is a positive integer.

4. Choose the most powerful test among the class of SPU tests.

$$\text{aSPU} = \min_{\gamma_1, \gamma_2} P_{\gamma_1, \gamma_2},$$

where P_{γ_1, γ_2} is the p-value of the SPU(γ_1, γ_2). According to our previous experience of applying the aSPU test, the results are not sensitive to the choice of Γ . $\Gamma_1 = \{1, 2, \dots, 8, \infty\}$ usually suffices in real-data application (Pan et al., 2014). Throughout the manuscript, we set $\Gamma_1 = \Gamma_2 = \{1, 2, \dots, 8, \infty\}$.

In $S(k, \gamma_1)$, $U_{k,j}^{\gamma_1}$ can be rewritten in an alternative form $U_{k,j}^{\gamma_1} = U_{k,j}^{\gamma_1-1} U_{k,j} = w_{k,j}^* U_{k,j} w_{k,j}^*$ can be considered as a type of data-driven weight for each score element because it is a function of $U_{k,j}$ which reflects the association strength and direction between RV j and the trait in study k . With $\gamma_1 = 1$, SPU test weights each RV equally and yields the highest power if all the RVs have similar effect sizes and the same association direction. When a subset of RVs are associated with the trait or their association directions are different, the SPU($\gamma_1 = 2, \gamma_2$) is often more powerful. As γ_1 increases, the SPU test puts heavier weights on the RVs which are more strongly associated with the trait. At the end, as the parameter approaches ∞ (as an even integer), it only considers the most significant RV, i.e., when $\gamma_1 = \infty$, $S(k, \gamma_1) = \max(|U_{k,j}|)$.

Similarly, γ_2 controls how much to up-weight the studies contain some associated RVs. When $\gamma_2 = 1$, studies are equally weighted and performs best when there is an almost equal association with the same direction across all studies. As γ_2 increases, SPU weights more on larger cohort-specific statistic. In an extreme case, SPU($\gamma_1, \gamma_2 = \infty$), one single cohort is considered, i.e., When $\gamma_2 = \infty$, $\text{SPU}(\gamma_1, \gamma_2) = \max(|S(k; \gamma_1)|)$. If one is interested in the most significant RV in a single study, SPU($\gamma_1 = \infty, \gamma_2 = \infty$) can be considered: $\text{SPU}(\gamma_1 = \infty, \gamma_2 = \infty) = \max |U_{k,j}|$. Using various combinations of γ_1 and γ_2 , one can target and fit different association patterns across multiple RVs and multiple studies, being doubly adaptive to both RVs and studies, thus achieving high power. Although not discussed here, our aSPU-meta test can easily accommodate external marker-specific and study-specific weights, which may improve power if they are correctly specified.

To assess the significance of each SPU(γ_1, γ_2) and aSPU-meta tests, we use a singlelayer of Monte-Carlo simulation. The Monte Carlo simulation method takes advantage of the closed-form formula of the variance of the score function, which has low computational burden. Specifically, the p-values of the SPU and aSPU-meta tests are obtained as follows.

Step 0. Obtain the $U_k^{(b)}$ under the null hypothesis using Monte Carlo simulations for $k = 1, \dots, K$ and $b = 1, \dots, B$, where b is the simulation iteration number and B is the total number of simulations. Specifically, if U_k and V_k are available from different studies, we simulate the score statistics under the null hypothesis: $U_k^{(b)} \sim \mathcal{MN}(0, V_k)$.

Step 1. Calculate the statistics $\text{SPU}(\gamma_1, \gamma_2)^{(b)}$ for $\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2$ as described.

Step 2. From the statistics $\text{SPU}(\gamma_1, \gamma_2)^{(b)}$, obtain the statistics $\text{aSPU}^{(b)}$ under null distribution:

$$P_{\gamma_1, \gamma_2}^{(b)} = \left[\sum_{b_1 \neq b} I(|\text{SPU}(\gamma_1, \gamma_2)^{(b_1)}| \geq |\text{SPU}(\gamma_1, \gamma_2)^{(b)}|) + 1 \right] / B,$$

$$\text{aSPU}^{(b)} = \min_{\gamma_1, \gamma_2} P_{\gamma_1, \gamma_2}^{(b)}.$$

Step 3. Based on the above statistics, the p-values of the $\text{SPU}(\gamma_1, \gamma_2)$ and aSPU are obtained:

$$P_{\text{SPU}(\gamma_1, \gamma_2)} = \left[\sum_{b=1}^B I(|\text{SPU}(\gamma_1, \gamma_2)^{(b)}| \geq |\text{SPU}(\gamma_1, \gamma_2)|) + 1 \right] / (B + 1),$$

$$P_{\text{aSPU-meta}} = \left[\sum_{b=1}^B I(|\text{aSPU}^{(b)}| \leq |\text{aSPU}|) + 1 \right] / (B + 1).$$

B is chosen to be 1,000 in the following simulation study. More generally, e.g. in real data applications, we can use a step-up procedure to determine B to save the computation time (Xu et al., 2017): we can increase B gradually until the genome-wide significance is achieved or until when it becomes obvious that the genome-wide significance is unlikely to be reached.

2.2 Relationship between the aSPU-meta and some existing tests

We compare our aSPU-meta test with some existing and popular methods, illustrating the advantage and novelty of our proposed method. Without loss of generality, we assume all of the marker-specific weights and between-study weights equal one. By some algebra, it can be shown that the meta-Burden test statistic is closely related to that of $\text{SPU}(1,1)$, and the test statistic of Het-meta-SKAT is equivalent to that of $\text{SPU}(2,1)$ (Lee et al., 2013), as follows.

$$\text{meta-Burden} = \left(\sum_{j=1}^p \sum_{k=1}^K U_{k,j} \right)^2 = \text{SPU}(1,1)^2,$$

$$\text{Het-meta-SKAT} = \sum_{j=1}^p \sum_{k=1}^K U_{k,j}^2 = \text{SPU}(2,1).$$

Furthermore, FE-Burden, RE-Burden, and FE-VC tests can be written as follows:

$$\begin{aligned}
 \text{FE-Burden} &= \left(\sum_{j=1}^p \sum_{k=1}^K U_{k,j} \right)^2 / \sum_{k=1}^K v_k = [\text{SPU}(1,1)]^2 / \nu_1, \\
 \text{RE-Burden} &= \text{FE-Burden} + \left(\sum_{j=1}^p \left[\sum_{k=1}^K U_{k,j} \right]^2 - \nu_1 \right)^2 / \left(2 \sum_{k=1}^K v_k^2 \right) = [\text{SPU}(1,1)]^2 / \nu_1 + [\text{SPU}(1,2) - \nu_1]^2 / \nu_2, \\
 \text{FE-VC} &= \sum_{j=1}^p \left(\sum_{k=1}^K U_{k,j} \right)^2 = \text{SPU}(2,1) \\
 &+ \text{CP},
 \end{aligned}$$

where $v_k = \frac{1}{\sigma_k^2} \left\{ \sum_{i=1}^{n_k} G_{ki} G'_{ki} - \left(\sum_{i=1}^{n_k} G_{ki} Z'_{ki} \right) \left(\sum_{i=1}^{n_k} G_{ki} Z'_{ki} \right)^{-1} \left(\sum_{i=1}^{n_k} Z_{ki} G'_{ki} \right) \right\}$ is the variance

of burden test, $G_{ki} = \sum_{j=1}^p X_{kij}$ is the genetic burden score, $\nu_1 = \sum_{k=1}^K v_k$ is the variance of SPU(1,1) test statistic, $\nu_2 = 2 \sum_{k=1}^K v_k^2$ is the variance of SPU(1,2) test statistic, and CP is a sum of some cross-products of $U_{k,j}$. Besides, RE-VC is a quadratic form of FE-VC and Het-meta-SKAT.

The above existing tests are closely related to the SPU(γ_1, γ_2) with γ_1 and γ_2 taking value one or two. Based on the previous study (Pan et al., 2014), we know that with more sparse association patterns across studies and/or across RVs, i.e., only few RVs in the given set are associated and/or RV associations are present only in few studies among all the anticipated studies in the meta-analysis, the existing tests might not perform well; larger values of γ_1 and γ_2 are needed to yield high power. In contrast, our aSPU-meta test combines multiple SPU tests with γ_1 and γ_2 ranging from small to large, thus it is able to maintain high power across a wide spectrum of scenarios.

Other hybrid/adaptive tests, such as SKATo or variant set mixed model association test (SMMAT-E), are also available for meta-analysis (Chen et al., 2019). These tests combine the burden and SKAT tests to better deal with different association patterns. In contrast, the aSPU-meta test combines many more SPU tests to account for more widely-ranging association patterns and between-study heterogeneities.

2.3 Data Availability

The NHLBI ESP data are accessible from the National Center for Biotechnology Information (NCBI) dbGap with accession numbers phs000401 (FHS), phs000398 (ARIC), phs000400 (CHS), phs000281 (WHI), and phs000402 (JHS). The study also makes use of data generated by the UK10K Consortium, derived from samples from the TwinsUK and ALSPAC cohorts. A full list of the investigators who contributed to the generation of the data is available from www.UK10K.org. Data are available from UK10K Data Access Committee for researchers who meet the criteria for access to confidential data. An R package aSPUmeta implementing the aSPU-meta test is publicly available at <https://github.com/ytzhong/metaRV> and will be on CRAN soon.

3 Results

3.1 Simulations

To evaluate and compare the proposed test with other existing methods, we considered different association patterns and various between-study heterogeneity levels. More specifically, we varied the numbers of causal variants and of the studies with causal variants, and included one or two racial/ethnic populations. We evaluated $K = 10$ studies, each composed of 1,000 unrelated subjects. The sequencing data were simulated using HAPGEN2 (Su et al., 2011) so that the simulated genotype data had the same linkage disequilibrium patterns as the target region of the 1000 Genomes reference. The average total exon length of a gene is known to be about 3 Kb (Lee et al., 2013; Pruitt et al., 2011); in each simulation, we generated $1,000 \times K$ haplotypes of length 4 Kb for a gene region *APOE* in chromosome 19, which is known to be associated with triglycerides (Peloso et al., 2014). Among the generated sequencing data, only variants with $MAF < 0.05$ were included. In study k , a quantitative phenotype was sampled from a Gaussian distribution with mean $\beta_k'X_k + 0.5$ and variance $\sigma_k^2 = 1$, where β_k was the vector of the true effect sizes for genotype matrix X_k .

In simulation set-up 1, we assumed that all studies came from the Utah Residents with Northern and Western European Ancestry population (CEU), such that the sequencing data for all studies were generated based on the CEU 1000 Genomes reference samples. A total of 15 RVs were included in the *APOE* region. Among 10 simulated studies, we randomly selected 1, 2, 5 or 7 studies carrying causal variants. For a study with causal variants, 2 or 7 of the 15 variants were randomly selected as causal. We allowed the causal variants to have their effect sizes varying across the studies: β_k was generated from $N(0.25, SD^2)$ for $k = 1, \dots, 10$, where we considered $SD \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, similar to that in (Tang and Lin, 2015).

In set-up 2, we further increased the between-study heterogeneity by including the Yoruba in Ibadan population (YRI). Out of $K = 10$ studies, we considered 1, 2, 5, or 7 studies composed of the CEU samples, and the rest were YRI samples. As before, the sequencing data were generated using HAPGEN2 for the *APOE* region. The simulated CEU samples included 15 RVs in the *APOE* region, while the simulated YRI samples included 30 RVs. In each simulation, we randomly selected 2 or 7 of variants generated from the CEU 1000 Genomes reference as causal variants; None of the variants generated from the YRI 1000 Genomes reference samples was selected as causal, mimicking the genetic heterogeneity observed in Pirim et al. (2019).

We compare the power of the aSPU-meta test with the existing methods based on 1000 replications, i.e., FE-VT, FE-VC RE-VT, and RE-VC tests in the MASS package (Tang and Lin, 2013, 2014), meta-Burden, Hom-meta-SKAT, Hom-meta-SKATo, Het-meta-SKAT and Het-meta-SKATo tests (Lee et al., 2013) in the MetaSKAT package on CRAN, and SMMAT-E in the GMMAT package on CRAN. Note that we used the default settings for the competing tests in their corresponding packages; that is, marker-specific beta weights were set to up-weight the rarer variants in the meta-Burden and Het(Hom)-meta tests.

Additionally, for SMMAT-E, two cohort groups were specified according to the simulated two populations in set-up 2, while one cohort group was specified to correspond to the only population in set-up 1; the studies within the same cohort group were assumed to have homogeneous effects.

To evaluate the Type I error rates of each method, we took out 100 genes located upstream and downstream of gene *APOE* on chromosome 19, and we ran 1E+4 replications for each gene under set-up 1 such that we were able to evaluate the aSPU-meta test at the nominal significance level $\alpha = 5E-5$ out of 1E+6 replications. The number of RVs per gene ranged from 5 to 160, with mean 52 and median 45.

The empirical Type I error rates of the aSPU-meta test were 4.84E-2, 5.01E-3, 4.39E-4, and 4.31E-5 at the nominal significance levels of 5E-2, 5E-3, 5E-4, 5E-4 and 5E-5 respectively, thus the Type 1 error rates were well controlled. Figure 1 presents the empirical power for seven methods. Table S1 shows the empirical power for the remaining methods, along with the top performers in Figure 1, under some simulation settings from set-ups 1 and 2. Clearly, the power of the tests increased with the effect sizes, the number of causal RVs, or the number of studies carrying the causal RVs, but decreased with a stronger study heterogeneity level. Some study heterogeneity was present in both simulation set-ups: in set-up 1, the causal variants were allowed to be different across the studies; in set-up 2, there was even stronger heterogeneity with the inclusion of the studies from a different population. Coinciding with the study heterogeneity levels, the tests based on the RE models (or similar ideas), such as aSPU-meta, Het-meta-SKAT, Het-meta-SKATo, RE-VT and RE-VC, generally had higher power than the tests based on the FE models, which was consistent with the previous findings (Han and Eskin, 2011; Lee et al., 2013; Tang and Lin, 2014). In set-up 1 (Figure 1, A and B), the power of the aSPU-meta test was close to that of Het-meta-SKAT and Het-meta-SKATo tests. In set-up 2 (Figure 1, C and D), the aSPU-meta test was the most powerful in most cases. Comparing the performance of the aSPU-meta test and the competing tests, we found that the aSPU-meta test showed its edge in the presence of stronger heterogeneity. Additionally, the aSPU-meta test outperformed other tests when the association signals were sparse. As expected, when the proportion of causal variants was small, a SPU test with a larger γ_1 had relatively high power. For example, when there were 7 studies with 2 causal variants, $SPU(\infty, 1)$ had larger power than $SPU(1, 1)$. When the proportion of studies carrying the causal variants was small, a SPU test with a larger γ_2 had relatively high power. For example, when there was only one study with 7 causal variants, $SPU(1, \infty)$ was more powerful than $SPU(\infty, 1)$. Therefore, γ_1 and γ_2 were confirmed to be effective tuning parameters for variant-specific association strengths and between-study heterogeneity levels respectively. In conclusion, the aSPU-meta test showed better performance than other methods under the scenarios when the signals were sparse in the region or heterogeneous among the studies. In other scenarios, even if the aSPU-meta test was not the most powerful one, its power was close to the winners'.

3.2 Analysis of the ESP data

To further demonstrate the performance of different meta-analysis methods for RVs, we analyzed the WES data in association with plasma triglyceride level in individuals of

European-ancestry (EA) and AA, who were sequenced in the NHLBI ESP project. Four cohorts for EA, i.e., Atherosclerosis Risk in Communities (ARIC), the Cardiovascular Heart Study (CHS), the Framingham Heart Study (FHS), and the Womens Health Initiative (WHI) were included, with a sample size of 503, 140, 417, and 671, respectively. Three cohorts for AA, i.e., ARIC, WHI and Jackson Heart Study (JHS) were included in the meta-analysis, with a sample size of 295, 735, and 310, respectively. We performed a natural logarithm transformation on the plasma triglyceride level and adjusted for covariates age, sex, and two principal components capturing possible population substructure (Crosby et al., 2014). We included nonsynonymous and splice-site variants of MAF $\geq 1\%$ within each gene and excluded genes with cumulative MACs < 5 , resulting in a total of 16,806 genes and the Bonferroni-adjusted genome-wide significance level at $3.0E-6$. The specific QC procedure was described previously (Crosby et al., 2014; Wei et al., 2016). For the aSPU-meta test, we calculated the p-values by using a stepwise procedure. We started with $B=1,000$ Monte Carlo simulations for all genes, and gradually increased B . If the estimated p-value was less than $10/B$, we increased B to $10 \times B$ until $B = 1E+7$ for the genome-wide significance.

There was no evidence of Type I error inflation for the aSPU-meta test as shown by a genomic control factor of 0.98 and a quantile-quantile plot (Figure S1); the results are also shown in a Manhattan plot (Figure S2). We identified five genes reaching the genome-wide significance in the meta-analysis of the seven studies: *APOC3*, *B4GALNT2*, *C19orf35*, *MYCT1*, and *USP12* (Table 1). *APOC3* and *MYCT1* were previously reported to be associated with plasma triglyceride level in EA (Crosby et al., 2014; Peloso et al., 2014; Timpson et al., 2014; Li et al., 2015) and AA (Crosby et al., 2014), respectively, while the other three genes are novel, and one of them was identified by the aSPU-meta test only. Other than *APOC3*, all signals were driven by AA. Table 2 provides the p-values of the performed tests for *APOC3* and *MYCT1*. *APOC3* was identified by the meta-Burden test. The SPU(1,1) test, which is equivalent to the meta-Burden test also reached genome-wide significance, and it was the most significant among all the SPU tests (Table S2). The p-value of the aSPU-meta test failed to reach the significance threshold, possibly due to the power loss by evaluating a number of tests at the same time. *MYCT1*, on the other hand, was identified by Hom(Het)-meta-SKAT(o) and aSPU-meta tests, among which the most significant test was Het-meta-SKATo. In addition, the SPU tests with $\gamma_1 = 1$ and $\gamma_2 = 1$ were overall more significant, suggesting a complex association pattern and strong between-study heterogeneity. Although aSPU-meta did not give the smallest p-value for *MYCT1*, it was still highly significant. By examining the results from the individual SPU tests, we might be able to get an understanding of the underlying genetic architecture and heterogeneity level. We made a closer examination of *B4GALNT2* in section 3.2.1 because it was only identified by the aSPU-meta test. *C19orf35* and *USP12* were identified by some other competing tests, but the aSPU-meta test yielded the smallest p-values.

3.2.1 A closer examination of the gene *B4GALNT2*—As demonstrated in Table 1 and Table 2, *B4GALNT2* was identified by a few SPU tests and the aSPU-meta test, but missed by other competing tests based on 7 studies. Among the competing tests, Het-meta-SKAT gave the most significant p-value of $3.1E-06$, close to the genome-wide significance level. To better evaluate the p-values of the SPU tests, we increased B to $1E+8$ (Table S3).

We found that the SPU tests with $\gamma_1 = 3$ and $\gamma_2 = 2$ were overall more significant than the other SPU tests, indicating that a large proportion of the 26 RVs might not be associated with the trait, and that the between-study heterogeneity level might be high.

One variant (chr17:47225736), mapped to the intergenic region of *B4GALNT2* with MAF < 0.01, was previously identified to be associated with coronary artery disease, a trait closely related with plasma triglyceride (Van der Harst and Verweij, 2017). This variant locates closely to the most significant RV (chr17:47219531) in the pooled single-variant analysis (Figure 2). More specifically, the WHI AA cohort made a major contribution to the significance of the pooled single-variant analysis with a p-value of 2.5E-7. Under this scenario, a SPU test with larger γ_1 and γ_2 was able to capture the signal, which explains why the p-values of the burden or VC based tests did not reach the genome-wide significance level. Since the two variants located at chr17:47219531 and chr17:47225736 are too rare, 1000 Genomes reference panel of the African population cannot provide a reliable estimation of their correlation, which warrants further investigation.

3.3 Validation using the UK10K data

The *APOC3* gene with a functional variant located at chr11:116701354 was previously identified to be associated with plasma triglyceride level among the UK10K cohorts by single-variant analysis (Timpson et al., 2014). We further examined the *APOC3* gene through our gene-based meta-analysis using two UK10K cohorts, i.e., the Avon Longitudinal Study of Parents and Children (ALSPAC) and TwinsUK cohorts. A total of 1,497 and 1,711 EA subjects with plasma triglyceride level measurements were included in the ALSPAC and TwinsUK, respectively. The genetic variants were measured using a WGS platform with an average $6.7 \times$ mean read-depth using Illumina NGS technology (More detailed description in (Timpson et al., 2014)). Since the lipid traits of the ALSPAC subjects were measured at age 9 years and subjects in the TwinsUK are all females, sex was adjusted in the ALSPAC while age and age-squared were adjusted in the TwinsUK cohort. By including RVs with MAF $\leq 1\%$, 1 Kb upstream and 1 Kb downstream of the *APOC3* gene, we had 29 RVs in the ALSPAC and 15 RVs in TWINS, with 10 overlapping RVs. The Hom-meta-SKAT test had the smallest p-value of 1.0E-4, while aSPU-meta followed with the second smallest p-value of 2.5E-3. Among all the SPU tests, the SPU(3,1) was the most significant test with a p-value of 5.2E-4, suggesting a quite sparse signal and a small amount of between-study heterogeneity across the two UK10K cohorts, whereas the p-value of SPU(2,1) test was 5.2E-3. Note that the Hom-meta-SKAT test is equivalent to the SPU(2,1), but the Hom-meta-SKAT used beta weights which favors rarer variants whereas the SPU test used equal weights. Therefore, we claimed that the additional power of the Hom-meta-SKAT test was mainly driven by the appropriate weighting in this case, but we noted that RVs with lower MAFs may not always have a higher impact on the phenotype of interest.

3.4 A combined analysis of the ESP and UK10K data

WES and WGS studies used different sequencing platforms. We expect a much stronger heterogeneity when combining WES and WGS studies. To evaluate the performance of the tests under more extremely high study- and variant-level heterogeneities, for illustration, we did a post hoc meta-analysis of the two UK10K and ESP cohorts. RVs with MAF $\leq 1\%$

were grouped by including RVs 1 Kb upstream and 1 Kb downstream of the gene in UK10K cohorts. Among a total of 9,270 genes with less than 250 RVs per gene, *APOC3* was still the only gene reached genome-wide significance. Two RVs were shared across UK10K and ESP studies. The VC tests of the *APOC3* gene were, in general, more powerful than burden tests, as opposed to their performance in the meta-analysis of ESP studies only, suggesting such combination of studies included different directions of the association between the grouped RVs and plasma triglyceride level. With more individuals and RVs included, the Hom-meta-SKAT and aSPU-meta tests became the top two performers with smaller p-values of $1.9\text{E-}9$ and $6.6\text{E-}6$, respectively. The most significant SPU test was SPU(3,1), which had a p-value of $9.0\text{E-}7$. It is interesting to note that the p-value of the aSPU-meta test was more significant than that in the ESP or UK10K study alone, suggesting a possible power gain by combining the two studies with two different types of sequencing data.

4 Discussion

In order to improve the statistical power of detecting RVs, lots of recent efforts have been put into developing powerful statistical tests (Derkach et al., 2013; Sha et al., 2013; Wang, 2016; Wei et al., 2016; Chen et al., 2017), leveraging information from multiple traits (Zhang et al., 2019), and incorporating gene-by-environment interaction (Zhao et al., 2015; Su et al., 2017; Yang et al., 2019). However, there is still limited findings. For CVs, most genetic risk variants discovered in the past few years came from large-scale meta-analyses of GWAS, indicating that a single study may be underpowered to detect CVs with moderate to small effect sizes (Zeggini and Ioannidis, 2009). It is thus natural to extend meta-analysis to sequencing studies because statistical power is, in general, a more serious problem for RVs (Auer and Lettre, 2015). In this paper, we have proposed a unified framework for meta-analysis of RVs, called the aSPU-meta test. Similar to the existing meta-analysis methods, it only requires sharing the summary statistics, instead of individual-level data, i.e., the score vector and its correlation matrix, which is considered a major advantage for collaborations across genetics studies. In addition, we showed that our aSPU-meta test was more powerful than many existing tests, being either the most powerful or close to the most powerful ones, across various simulation set-ups, and in particular, was able to identify novel genes in our real data analysis.

The power of meta-analysis for RVs not only depends on the genetic architecture but also the between-study heterogeneity. For the former, it is known that there is no uniformly most powerful test for RV set association testing (Pan, 2009). For the latter, we may expect strong between-study heterogeneity levels when multiple ethnic groups are included or different sequencing platforms are used. In the real-data example, we explicitly examined the combination of two populations, EA and AA. With the inclusion of AA, we found four more genes associated with plasma triglyceride level than using EA alone, which is in agreement with the previous observation that study populations including AA had more power to detect association with RVs (Mensah-Ablorh et al., 2016). In addition, we examined a combined meta-analysis of a WES dataset and a WGS dataset, where we found a different set of best-performing tests and showed benefit of increased power when the heterogeneities were well accommodated. Furthermore, we remark that the genetic architecture and the between-study heterogeneity vary from gene to gene while their true values are unlikely to be known.

Therefore, it is crucial to run a test that is powerful across a wide range of scenarios. However, existing tests are usually powerful under only limited scenarios. Our aSPU-meta test fills in this gap, summarizing the information across a large number and variety of tests, i.e., 9×9 SPU tests in our analyses, to achieve high power. In our data example, gene *B4GALNT2* was discovered by the aSPU-meta test for the AA subjects in the presence of high between-study heterogeneity and a large number of likely neutral RVs; in contrast, it was missed by competing methods. Although we did not focus on the use of variant-specific or study-specific weights, appropriate incorporation of such weights may increase the statistical power and interpretation, for example, using the functional weights (Zhan and Liu, 2015; He et al., 2017; Ma and Wei, 2019). We also note that, similar to other competing tests for RVs, one potential drawback of the current implementation of the aSPU-meta test using Monte Carlo simulations is its use of the asymptotic normal distribution of the score vector, which may not hold for extremely unbalanced case-control studies, and other resampling methods like permutation-based ones may need to be adopted for more accurate p-value calculations (Dey et al., 2019).

Lastly, we remark that it is computationally feasible to run the aSPU-meta test with only one layer of Monte Carlo simulations. Time for computing the p-value of the aSPU-meta test depends on the genome-wide significance level, the number of RVs, and the number of studies. For example, it took around 3 minutes to run a total of 10 million Monte Carlo simulations for *APOC3* gene with 39 RVs in nine study cohorts (two UK10K cohorts and seven ESP cohorts). With the help of parallel computing and stepwise procedure, the computation time of the whole genome scan can be further reduced. In the application of our meta-analysis on 16,806 genes from the seven ESP cohorts, it took 0.3 hours on average for using 528 computing cores with 2.5GB memory limit to compute the p-values. Further improvement is possible: a recent implementation of the aSPU-meta test in Julia has made it possible to run up to $B = 1E + 11$ simulations to yield highly significant p-values (Baldassari et al., 2019). We provide an R package aSPUmeta implementing the aSPU-meta test to facilitate its use.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the reviewers for many helpful comments and suggestions. This work was supported by the National Institutes of Health grants R21AG057038, R01HL116720, R01GM113250, R01GM126002 and R01HL105397. We thank the Minnesota Supercomputing Institute and Texas Advanced Computing Center for providing computing resources.

References

- Auer PL and Lettre G (2015). Rare variant association studies: considerations, challenges and opportunities. *Genome Medicine*, 7(1):16. [PubMed: 25709717]
- Auer PL, Teumer A, Schick U, O'Shaughnessy A, Lo KS, Chami N, Carlson C, De Denus S, Dubé MP, Haessler J, et al. (2014). Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nature Genetics*, 46(6):629–634. [PubMed: 24777453]

- Baldassari AR, Sitlani CM, Highland HM, Arking DE, Buyske S, Darbar D, Gondalia R, Graff M, Guo X, Heckbert SR, et al. (2019). Multi-ethnic genome-wide association study of decomposed cardioelectric phenotypes illustrates strategies to identify and characterize evidence of shared genetic effects for complex traits. *BioRxiv*, page 654012.
- Chen H, Huffman JE, Brody JA, Wang C, Lee S, Li Z, Gogarten SM, Sofer T, Bielak LF, Bis JC, et al. (2019). Efficient variant set mixed model association tests for continuous and binary traits in large-scale whole-genome sequencing studies. *The American Journal of Human Genetics*, 104(2):260–274. [PubMed: 30639324]
- Chen S and Parmigiani G (2007). Meta-analysis of BRCA1 and BRCA2 penetrance. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 25(11):1329. [PubMed: 17416853]
- Chen Z, Lin T, and Wang K (2017). A powerful variant-set association test based on chi-square distribution. *Genetics*, 207(3):903–910. [PubMed: 28912342]
- Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, and Hobbs HH (2006). Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proceedings of the National Academy of Sciences*, 103(6):1810–1815.
- Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitzel NO, Lange LA, Lu Y, Tang ZZ, Zhang H, Hindy G, Masca N, et al. (2014). Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *The New England Journal of Medicine*, 371(1):22. [PubMed: 24941081]
- Derkach A, Lawless JF, and Sun L (2013). 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):110–121. [PubMed: 23032573]
- Dey R, Nielsen JB, Fritsche LG, Zhou W, Zhu H, Willer CJ, and Lee S (2019). Robust meta-analysis of biobank-based genome-wide association studies with unbalanced binary phenotypes. *Genetic epidemiology*, 43(5):462–476. [PubMed: 30793809]
- Fan R, Wang Y, Boehnke M, Chen W, Li Y, Ren H, Lobach I, and Xiong M (2015). Gene level meta-analysis of quantitative traits by functional linear models. *Genetics*, 200(4):1089–1104. [PubMed: 26058849]
- Fan R, Wang Y, Chiu C. y., Chen W, Ren H, Li Y, Boehnke M, Amos CI, Moore JH, and Xiong M (2016). Meta-analysis of complex diseases at gene level with generalized functional linear models. *Genetics*, 202(2):457–470. [PubMed: 26715663]
- Gibson G (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics*, 13(2):135–145.
- Han B and Eskin E (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Journal of Human Genetics*, 88(5):586–598. [PubMed: 21565292]
- Han B and Eskin E (2012). Interpreting meta-analyses of genome-wide association studies. *PLoS Genetics*, 8(3):e1002555. [PubMed: 22396665]
- He Q, Zhang HH, Avery CL, and Lin D (2016). Sparse meta-analysis with high-dimensional data. *Biostatistics*, 17(2):205–220. [PubMed: 26395907]
- He Z, Xu B, Lee S, and Ionita-Laza I (2017). Unified sequence-based association tests allowing for multiple functional annotations and meta-analysis of noncoding variation in metabochip data. *The American Journal of Human Genetics*, 101(3):340–352. [PubMed: 28844485]
- Lee S, Abecasis GR, Boehnke M, and Lin X (2014). Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics*, 95(1):5–23. [PubMed: 24995866]
- Lee S, Teslovich TM, Boehnke M, and Lin X (2013). General framework for meta-analysis of rare variants in sequencing association studies. *The American Journal of Human Genetics*, 93(1):42–53. [PubMed: 23768515]
- Li A, Morrison A, Kovar C, Cupples A, Brody J, Polfus LM, Yu B, Metcalf G, Muzny D, Veeraraghavan N, et al. (2015). Analysis of loss-of-function variants and 20 risk factor phenotypes in 8,554 individuals identifies loci influencing chronic disease. *Nature Genetics*, 47(6):640. [PubMed: 25915599]

- Liu D, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, Nikpay M, Auer PL, Goel A, Zhang H, et al. (2014). Meta-analysis of gene-level tests for rare variant association. *Nature Genetics*, 46(2):200. [PubMed: 24336170]
- Liu DJ and Leal SM (2012a). Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. *The American Journal of Human Genetics*, 91(4):585–596. [PubMed: 23022102]
- Liu DJ and Leal SM (2012b). A unified method for detecting secondary trait associations with rare variants: application to sequence data. *PLoS Genetics*, 8(11):e1003075. [PubMed: 23166519]
- Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam D, Alves AC, et al. (2017). Exome-wide association study of plasma lipids in > 300,000 individuals. *Nature Genetics*, 49(12):1758. [PubMed: 29083408]
- Lu X, Peloso GM, Liu DJ, Wu Y, Zhang H, Zhou W, Li J, Tang C. S. -m., Dorajoo R, Li H, et al. (2017). Exome chip meta-analysis identifies novel loci and east asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nature Genetics*, 49(12):1722. [PubMed: 29083407]
- Lumley T, Brody J, Dupuis J, and Cupples A (2012). Meta-analysis of a rare-variant association test Technical report, Technical report, University of Auckland <http://stattech.wordpress.fos.auckland.ac.nz/files/2012/11/skat-meta-paper.pdf>.
- Ma Y and Wei P (2019). Funspu: a versatile and adaptive multiple functional annotation-based association test of whole-genome sequencing data. *PLoS genetics*, 15(4):e1008081. [PubMed: 31034468]
- MacArthur D, Manolio T, Dimmock D, Rehm H, Shendure J, Abecasis G, Adams D, Altman R, Antonarakis S, Ashley E, et al. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature*, 508(7497):469. [PubMed: 24759409]
- Mensah-Ablorh A, Lindstrom S, Haiman CA, Henderson BE, Marchand LL, Lee S, Stram DO, Eliassen AH, Price A, and Kraft P (2016). Meta-analysis of rare variant association tests in multiethnic populations. *Genetic Epidemiology*, 40(1):57–65. [PubMed: 26639010]
- Morrison AC, Huang Z, Yu B, Metcalf G, Liu X, Ballantyne C, Coresh J, Yu F, Muzny D, Feofanova E, et al. (2017). Practical approaches for whole-genome sequence analysis of heart-and blood-related traits. *The American Journal of Human Genetics*, 100(2):205–215. [PubMed: 28089252]
- Pan W (2009). Asymptotic tests of association with multiple snps in linkage disequilibrium. *Genetic Epidemiology*, 33(6):497–507. [PubMed: 19170135]
- Pan W, Kim J, Zhang Y, Shen X, and Wei P (2014). A powerful and adaptive association test for rare variants. *Genetics*, 197(4):1081–1095. [PubMed: 24831820]
- Pan W, Kwak I-Y, and Wei P (2015). A powerful pathway-based adaptive test for genetic association with common or rare variants. *The American Journal of Human Genetics*, 97(1):86–98. [PubMed: 26119817]
- Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitzel NO, Brody JA, Khetarpal SA, Crosby JR, Fornage M, et al. (2014). Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *The American Journal of Human Genetics*, 94(2):223–232. [PubMed: 24507774]
- Pirim D, Radwan ZH, Wang X, Niemsiri V, Hokanson JE, Hamman RF, Fein-gold E, Bunker CH, Demirci FY, and Kamboh MI (2019). Apolipoprotein e-c1-c4-c2 gene cluster region and inter-individual variation in plasma lipoprotein levels: a comprehensive genetic association study in two ethnic groups. *PloS one*, 14(3):e0214060. [PubMed: 30913229]
- Pruitt KD, Tatusova T, Brown GR, and Maglott DR (2011). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research*, 40(D1):D130–D135. [PubMed: 22121212]
- Sha Q, Wang S, and Zhang S (2013). Adaptive clustering and adaptive weighting methods to detect disease associated rare variants. *European Journal of Human Genetics*, 21(3):332. [PubMed: 22781093]
- Su Y, Di C, and Hsu L (2017). A unified powerful set-based test for sequencing data analysis of GxE interactions. *Biostatistics*, 18(1):119–131. [PubMed: 27474101]

- Su Z, Marchini J, and Donnelly P (2011). HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics*, 27(16):2304–2305. [PubMed: 21653516]
- Tang Z and Lin D (2013). MASS: meta-analysis of score statistics for sequencing studies. *Bioinformatics*, 29(14):1803–1805. [PubMed: 23698861]
- Tang Z and Lin D (2014). Meta-analysis of sequencing studies with heterogeneous genetic associations. *Genetic Epidemiology*, 38(5):389–401. [PubMed: 24799183]
- Tang Z and Lin D (2015). Meta-analysis for discovering rare-variant associations: statistical methods and software programs. *The American Journal of Human Genetics*, 97(1):35–53. [PubMed: 26094574]
- Timpson NJ, Walter K, Min JL, Tachmazidou I, Malerba G, Shin S-Y, Chen L, Futema M, Southam L, Iotchkova V, et al. (2014). A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in europeans. *Nature Communications*, 5:4871.
- Van der Harst P and Verweij N (2017). The identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circulation Research*, 122(3):433–443. [PubMed: 29212778]
- Wang K (2016). Boosting the power of the sequence kernel association test by properly estimating its null distribution. *The American Journal of Human Genetics*, 99(1):104–114. [PubMed: 27292111]
- Wei P, Cao Y, Zhang Y, Xu Z, Kwak I-Y, Boerwinkle E, and Pan W (2016). On robust association testing for quantitative traits and rare variants. *G3: Genes, Genomes, Genetics*, 6(12):3941–3950. [PubMed: 27678522]
- Xu Z, Wu C, Wei P, and Pan W (2017). A powerful framework for integrating eqtl and gwas summary data. *Genetics*, 207(3):893–902. [PubMed: 28893853]
- Yang T, Chen H, Tang H, Li D, and Wei P (2019). A powerful and data-adaptive test for rare-variant-based gene-environment interaction analysis. *Statistics in Medicine*, 38(7):1230–1244. [PubMed: 30460711]
- Yano K, Yamamoto E, Aya K, Takeuchi H, Lo P, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K, et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature Genetics*, 48(8):927–934. [PubMed: 27322545]
- Zeggini E and Ioannidis JP (2009). Meta-analysis in genome-wide association studies. *Pharmacogeomics*, 10(2):191–201.
- Zhan X and Liu DJ (2015). SEQMINER: An R-package to facilitate the functional interpretation of sequence-based associations. *Genetic Epidemiology*, 39(8):619–623. [PubMed: 26394715]
- Zhang J, Sha Q, Liu G, and Wang X (2019). A gene-based approach to test genetic association based on an optimally weighted combination of multiple traits. *bioRxiv*, page 592568.
- Zhao G, Marceau R, Zhang D, and Tzeng J-Y (2015). Assessing gene-environment interactions for common and rare variants with binary traits using gene-trait similarity regression. *Genetics*, 199(3):695–710. [PubMed: 25585620]

Figure 1:

Empirical power curves under simulation set-up 1 (A and B) and set-up 2 (C and D). A and C: 2 of the included variants are causal. B and D: 7 of the included variants are causal. # cstudy is the number of the studies with causal variants out of the total 10 studies. SD is the standard deviation of the effect sizes for the causal variants.

Figure 2:
Locus Zoom plot of the gene *B4GALNT2* based on the p-values of the pooled single-variant analysis from the three studies with AA subjects (build: hg19)

Table 1:

Genome-wide significant genes from the meta-analysis of 7 ESP studies. The contributing population (CP) is determined by performing meta-analysis on EA and AA separately. The RVs with p-values of the pooled single-variant analysis < 0.05 are marked out. SKAT(o) refers to both SKAT and SKATo; similarly for Hom(Het).

Gene	Full Gene Name	Significant tests	#of RVs	CP	RV location in the CP (build: hg19)
<i>APOC3</i>	apolipoprotein C-III	meta-Burden, SPU(1,1)	7	EA	chr11:116701284, chr11:116701353 , chr11:116701354 , chr11:116701560 , chr11:116701613, chr11:116703493
<i>B4GALNT2</i>	Beta-1,4-N-Acetyl-Galactosaminyl Transferase 2	aSPU-meta, SPU($\gamma_1 = 2, \gamma_2 = 2$)	26	AA	chr17:47218665, chr17:47219419, chr17:47219528, chr17:47219531 , chr17:47230176, chr17:47233954, chr17:47241514, chr17:47241533, chr17:47241625, chr17:47243608, chr17:47246074, chr17:47246239, chr17:47246242 , chr17:47246894, chr17:47246956, chr17:47247013, chr17:47247014, chr17:47247085
<i>C19orf35</i>	Chromosome 19 Open Reading Frame 35	aSPU-meta, SPU($\gamma_1 = 2, \gamma_2 = 2$)Het-meta-SKAT(o)	5	AA	chr19:2280863 , chr19:2280903, chr19:2280920
<i>MYCT1</i>	MYC Target 1	aSPU-meta, SPU($\gamma_1 = 2, \gamma_2 = 2$) Hom(Het)-meta-SKAT(o)	16	AA	chr6:153019189, chr6:153019197 , chr6:153042982, chr6:153042987, chr6:153043077, chr6:153043263, chr6:153043291, chr6:153043303 , chr6:153043311
<i>USP12</i>	Ubiquitin Specific Peptidase 12	aSPU-meta, SPU($\gamma_1 = 2, \gamma_2 = 2$) Het-meta-SKAT(o)	4	AA	chr13:27669856 , chr13:27690726

Table 2:

Three genome-wide significant genes in the meta analysis of 7 ESP studies. The p-values smaller than the genome-wide significance threshold are marked in boldface. The p-values of the SPU and aSPU-meta tests were based on 1E+7 Monte Carlo simulations.

Tests	<i>APOC3</i>			<i>MYCT1</i>			<i>B4GALNT2</i>		
	ALL	EA	AA	ALL	EA	AA	ALL	EA	AA
meta-Burden	1.7E-6	6.7E-6	2.3E-2	7.7E-5	5.4E-1	1.4E-5	9.3E-01	6.1E-01	8.7E-01
Hom-meta-SKAT	6.0E-5	5.5E-5	1.3E-1	2.9E-7	8.6E-1	7.4E-9	2.1E-04	6.4E-01	1.8E-04
Het-meta-SKAT	8.5E-4	5.8E-4	1.0E-1	2.9E-8	8.6E-1	3.0E-10	3.1E-06	5.5E-01	2.63E-06
Hom-meta-SKATo	3.9E-6	5.5E-6	3.8E-2	8.2E-7	7.4E-1	1.3E-8	5.0E-04	8.2E-01	4.4E-04
Het-meta-SKATo	4.6E-6	1.2E-5	3.6E-2	9.6E-10	7.6E-1	7.0E-10	1.2E-05	7.6E-01	9.92E-06
FE-VT	1.1E-4	6.9E-5	9.7E-3	6.9E-5	5.2E-1	5.1E-5	1.7E-01	6.8E-01	1.1E-01
RE-VT	1.0E-4	7.0E-5	9.1E-3	5.2E-5	5.0E-1	5.0E-5	1.3E-01	7.1E-01	1.4E-01
SPU(1,1)	6.0E-7	1.9E-6	1.6E-2	8.6E-5	5.7E-1	1.4E-5	8.3E-01	5.7E-01	9.8E-01
SPU(2,2)	8.5E-4	5.2E-4	1.2E-1	1.0E-7	8.7E-1	1.0E-7	2.0E-06	5.6E-01	1.7E-06
SPU(1,∞)	5.1E-3	8.7E-4	1.0E-2	1.2E-05	4.2E-1	9.5E-6	1.5E-01	7.1E-01	1.5E-01
SPU(∞,1)	1.4E-2	1.4E-3	4.9E-1	6.9E-06	9.1E-1	1.0E-7	3.8E-05	5.9E-01	5.6E-06
SPU(∞,∞)	4.5E-2	1.6E-2	1.8E-1	1.0E-7	9.1E-1	1.0E-7	1.0E-07	6.6E-01	2.0E-07
aSPU-meta	3.6E-6	9.6E-06	4.7E-2	1.0E-7	7.7E-1	1.0E-7	8.0E-07	8.0E-01	1.4E-06