Published in final edited form as:

Genet Epidemiol. 2016 November ; 40(7): 534–543. doi:10.1002/gepi.21996.

A Meta-Analytic Framework for Detection of Genetic Interactions

Yulun Liu^{1,2}, Yong Chen^{3,*}, and Paul Scheet^{2,1,4,*}

¹Division of Biostatistics, The University of Texas School of Public Health, Houston, TX 77030, USA

²Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

³Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, USA

⁴The University of Texas Graduate School of Biomedical Sciences, Houston, TX 77030, USA

Abstract

With varying, but substantial, proportions of heritability remaining unexplained by summaries of single-SNP genetic variation, there is a demand for methods that extract maximal information from genetic association studies. One source of variation that is difficult to assess is genetic interactions. A major challenge for naive detection methods is the large number of possible combinations, with a requisite need to correct for multiple testing. Assumptions of large marginal effects, to reduce the search space, may be restrictive and miss higher-order interactions with modest marginal effects. In this paper, we propose a new procedure for detecting gene-by-gene interactions through heterogeneity in estimated low-order (e.g. marginal) effect sizes by leveraging population structure, or ancestral differences, among studies in which the same phenotypes were measured. We implement this approach in a meta-analytic framework, which offers numerous advantages, such as robustness and computational efficiency, and is necessary when data-sharing limitations restrict joint analysis. We effectively apply a dimension reduction procedure that scales to allow searches for higher-order interactions. For comparison to our method, which we term phylogen Y-aware Effect-size Tests for Interactions (YETI), we adapt an existing method that assumes interacting loci will exhibit strong marginal effects to our meta-analytic framework. As expected, YETI excels when multiple studies are from highly differentiated populations and maintains its superiority in these conditions even when marginal effects are small. When these conditions are less extreme, the advantage of our method wanes. We assess the Type-I error and power characteristics of complementary approaches to evaluate their strengths and limitations.

Keywords

case-control design; gene-gene interaction; heterogeneity; mega-analysis; meta-analysis

*Correspondence to: Paul Scheet, Dept. of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. pscheet@alum.wustl.edu or Yong Chen, Dept. of Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania. ychen123@mail.med.upenn.edu.

The authors declare to have no conflicts of interest.

Introduction

Genome-wide association (GWA) studies have been widely used to explore associations between genetic variants and complex diseases such as diabetes and inflammatory bowel disease (Duerr *et al.*, 2006; Sladek *et al.*, 2007; Saxena *et al.*, 2007), where the genetic influences on phenotypes are polygenic. The standard GWA study has focused on identifying individual single nucleotide polymorphisms (SNPs) that are associated with complex traits or diseases. However, existing findings from the standard GWA studies can explain only a modest proportion of heritability for complex traits (Manolio *et al.*, 2009; Eichler *et al.*, 2010). Many complex traits are regulated by biological mechanisms that may involve multiple interacting genes. Not surprisingly, it has been suggested that incorporating gene-gene interactions can not only help to further explain the missing heritability of complex traits, but also can further characterize the genetic architecture of traits (Moore & Williams, 2009). However, there are at least two main challenges in identification of novel gene-gene interactions. The first is the computational burden; with a million SNPs, there are $\approx 10^{11}$ possible pairs of SNPs to be tested, just for 2-way interactions. The second (related) challenge is the limited power due to the requisite multiple testing correction.

To tackle the computational challenge of exhaustive searching, several machine-learning and data-mining approaches have been developed, such as multifactor dimensionality reduction for identifying high-order gene-gene interactions in association studies (Ritchie *et al.*, 2001, 2003). Some filtering techniques have also been used, such as the ReliefF algorithm (Kononenko, 1994) and TuRF (Moore & White, 2007).

The decision tree-based methods have been developed for large sample sizes, including random forest algorithms (Lunetta *et al.*, 2004; Bureau *et al.*, 2005; Jiang *et al.*, 2009). Statistical approaches have also been developed and include regression-based methods, such as Tukey's one-degree-of-freedom test (Chatterjee *et al.*, 2006), penalized logistic regression (Tanck *et al.*, 2006; Park & Hastie, 2008), two-stage analyses (Evans *et al.*, 2006; Kooperberg & LeBlanc, 2008; Murcray *et al.*, 2009), among others. For reviews, see Cordell (2009), Van Steen (2011) and Wei *et al.* (2014). However, all aforementioned methods are based on data from a single study and are based on the assumption that the individual level data are available. In this paper, we consider the situation where multiple studies are considered but only summary statistics are available for sharing among studies.

Recently, combining information across multiple GWA studies has become nearly standard practice, providing greater statistical power and providing validated conclusions (McCarthy *et al.*, 2008; Guerra & Goldstein, 2009; Begum *et al.*, 2012; Evangelou & Ioannidis, 2013). This "community" approach has major advantages over single study analyses because larger sample sizes allow investigation of rare variants (e.g., frequencies lower than 1%) (Cantor *et al.*, 2010) and provide great power to identify SNPs with small effect sizes (Guerra & Goldstein, 2009; Begum *et al.*, 2012). When individual level data from all studies are available, one can jointly analyze the pooled data, known as mega-analysis. However, this may be impractical for several reasons. First, obtaining individual level data may be limited by the terms of the informed consent within each study. Second, integration of very large genetic and phenotypic data sets from different studies is time-consuming and poses

additional challenges of data management, storage, and harmonization. Third, meta-analysis allows for analyses of individual studies to account for local population substructure, relationships among subjects, study-specific covariates, and other ascertainment related issues that may be optimally considered within each study. Conveniently, meta-analysis can be conducted where summary statistics from multiple studies are integrated together. There are generally two types of meta-analytic approaches. The first is to combine the p -values from multiple studies directly, and the second is to model the effects sizes from the combined studies (Guerra & Goldstein, 2009; Song & Tseng, 2014). Throughout this paper, we focus on the latter method in combining multiple studies, as this is typical in most meta-analyses of GWA studies to look for statistically significant single-SNP effects.

The goal of this paper is twofold. One, we seek to formalize statistical procedures to detect gene-gene interactions using data from multiple studies, relevant in this era of genetic consortia. Specifically, we extend the two-step testing procedure proposed by Kooperberg & LeBlanc (2008) (hereafter referred to as KL) from a single study to multiple studies (hereafter referred to as KL-meta). We show that analysis can be done purely based on summary statistics, avoiding the practical limitations and feasibility of pooling individual level data. Two, we introduce a novel framework to detect genetic interactions using heterogeneity in marginal effects across studies as a signal of an additional genetic factor, which could be an interacting locus, unaccounted for in the statistical model. We then subsequently test for interactions directly in a full statistical model via a genome-wide scan for every candidate marker identified. Our method, which we term phylogenY-aware Effect-size Tests for Interactions (YETI), will have power when there is sufficient population structure, or differences in allele frequencies across studies, such that unmeasured differential exposures of interacting loci induce changes in estimated marginal effects. These unmeasured variants affecting phenotype are considered a nuisance in obtaining accurate effect size estimates; here, we leverage this heterogeneity in its own test.

The major advantages of YETI, compared to KL-meta, are the following: first, unlike KL, the power function of YETI is monotonic in the interaction effect; second, it is indeed possible biologically that marginal effects are small but the interaction is not (Culverhouse *et al.*, 2002), and YETI is motivated especially for this. Hence, YETI may have superior power to KL-meta under certain scenarios. However, regardless of any pure advantages, the tests are complementary, since they leverage orthogonal pieces of information.

The remainder of this paper is organized as follows. We describe the extension to KL and our effect size heterogeneity test (YETI) in Methods. We also show that both KL-meta and YETI enjoy the filtering property (Dai *et al.*, 2012) of the two-step procedure. In Results, we compare the power of the proposed procedures under several simulated settings, varying population differentiation and marginal effect sizes, and discuss the strengths and limitations of the proposed procedures. Finally, in the Discussion, we briefly discuss our results and possible extensions to these methods.

Methods

Model Setting for Detection of G×G Interactions with Multiple Studies

Consider a collection of K association studies, each with subjects sampled from individual homogeneous populations arbitrarily divergent from each other; for example, each study may be comprised of individuals with either European, African or East Asian ancestries (but of minimal within-study heterogeneity). Let $Y_i^{(k)}$ be the disease status of the i -th subject in the k -th population and $G_i^{(k)} = (G_{i1}^{(k)}, \dots, G_{iL}^{(k)})$ be the genotypes at L loci, which are typed in all K studies. For simplicity, we consider a dominant genetic model and thus $G_{il}^{(k)}$ may take values 0 for two copies of a putative non-risk allele (in practice, may use major allele), or 1 otherwise. Two SNPs, say j and l , possibly interact to affect phenotype as follows:

$$\text{logit} \left\{ \Pr \left(Y_i^{(k)} = 1 | G_{ij}^{(k)}, G_{il}^{(k)} \right) \right\} = \beta_0^{(k)} + \beta_1^{(k)} G_{ij}^{(k)} + \beta_2^{(k)} G_{il}^{(k)} + \beta_3^{(k)} G_{ij}^{(k)} \times G_{il}^{(k)}, \quad (1)$$

where $\beta_0^{(k)}$ is the prevalence (in logit scale) of disease among noncarriers at SNPs j and l in the k -th study population, $\beta_1^{(k)}$ quantifies the main effect of SNP j among noncarriers at SNP l , $\beta_2^{(k)}$ quantifies the main effect of SNP l among noncarriers at SNP j , and $\beta_3^{(k)}$ is the interaction effect between SNPs j and l . Here the *main* effects $\beta_1^{(k)}$ and $\beta_2^{(k)}$ need to be differentiated from the *marginal* effects, $\eta_1^{(k)}$ and $\eta_2^{(k)}$ defined in equation (3) below, where the other SNP and the interaction are not adjusted. We note that we are not asserting differences in main effects across populations but expect estimated marginal effects could vary per the motivation for our framework.

To borrow information across multiple studies, we may test for an overall interaction effect, denoted as β_3^{overall} , which can be estimated by a univariate fixed-effect or random-effect meta-analysis procedure (DerSimonian & Laird, 1986). Specifically, let $\hat{\beta}_3^{(k)}$ denote the estimated interaction effect in the k -th study, and $\text{var}(\hat{\beta}_3^{(k)})$ denote its estimated variance. Under a fixed-effect model, β_3^{overall} can be estimated by

$$\hat{\beta}_3^{\text{overall}} = \frac{\sum_{k=1}^K w_k \hat{\beta}_3^{(k)}}{\sum_{k=1}^K w_k}, \quad (2)$$

where $w_k = 1/\text{var}(\hat{\beta}_3^{(k)})$. The variance of $\hat{\beta}_3^{\text{overall}}$ can be estimated by $1/\sum_{k=1}^K w_k$. A Wald test for interaction between two SNPs, aggregating information across studies, can be performed by testing $H_0: \beta_3^{\text{overall}} = 0$. A similar random-effect meta-analysis can be conducted by using a different weight, i.e. $1/\left\{ \hat{\tau}^2 + \text{var}(\hat{\beta}_3^{(k)}) \right\}$, where $\hat{\tau}^2$ is an estimate of between-study heterogeneity (DerSimonian & Laird, 1986).

To deal with the computational and multiple-testing burdens that come with a naive search of all pairwise interactions, approaches that limit the hypotheses by selecting a subset of markers for testing have been proposed. The KL method assumes interacting loci will exhibit larger marginal effects. Intuitively, this is appealing; SNPs with significant marginal effects are at least more likely to be functional in some sense and thus make attractive candidates for interactions *a priori*. Where the KL method leverages expected changes in *mean* phenotype given differential genetic exposures, YETI is based on *variation* in estimated effect sizes. Like KL, our procedure consists of two steps – a filtering step to narrow down candidate SNPs and a second step where these candidates are tested directly for interactions. Our method can be used to address the following two limitations acknowledged by Kooperberg & LeBlanc (2008) implicitly or explicitly: 1) KL is not powered for interactions involving loci with small or null *marginal* effects; and 2) KL may exhibit a non-monotonic power function, when marginal and interaction effects exert in opposite directions. YETI is a natural complement to KL and increasingly attractive when there exists genetic structure that can be easily identified and modeled, such as in consortia of studies from different populations.

Meta-Analytic Statistical Detection of Interactions

The first procedure we describe below is an extension of KL from a single study to multiple studies, which screens the candidate SNPs for interaction test by first evaluating their pooled marginal effects. By a simple argument, such a procedure is shown to inherit the independent filtering property of KL for a single study (Dai *et al.*, 2012). As noted in Kooperberg & LeBlanc (2008), although their method has superior power and better computational efficiency than a naive “all pairs” method, it may lead to non-monotonic power functions of interaction effect size. To circumvent such a problem and to harness additional information in the presence of structured populations, we propose a novel procedure based on the heterogeneity across studies of effect size estimates. The intuition is that *a substantial heterogeneity in marginal effects at a SNP across studies may suggest interactions between this SNP and the SNPs with differential allele frequencies among populations*. Once heterogeneity at a SNP is declared at a certain significance level, interactions between this SNP and all other SNPs can be tested. Below, we describe these two procedures in detail.

An Extension of KL to Multiple Studies—We first describe the following adaptation of KL to accommodate data from multiple studies *using summary statistics only*:

1. Step 1(a): *Subgroup analysis, conducted within each study*

For the k -th population, we fit the following logistic regression model at the l -th marker

$$\text{logit} \left\{ \Pr \left(Y_i^{(k)} = 1 \right) \right\} = \eta_{0(l)}^{(k)} + \eta_{1(l)}^{(k)} G_{il}^{(k)}, \quad (3)$$

for $l = 1, \dots, L$ and $k = 1, \dots, K$. Here $\eta_{1(l)}^{(k)}$ is the *marginal effect* of SNP l in the k -th population. In this step, covariates specific to the study or individual may be added to reduce residual variation or adjust for confounding influences on phenotype.

2. Step 1(b): *Screening SNPs with significant overall marginal effects, using summary statistics*

Following the idea of Kooperberg & LeBlanc (2008), we screen candidate SNPs with significant overall marginal effects before conducting tests of interaction. Standard univariate meta-analysis procedures can be used to

obtain overall marginal effects by combining estimates $\hat{\eta}_{1(l)}^{(k)}$ from K studies. Specifically, at this step, we test for $H_{0l}: \eta_{1(l)}^{\text{overall}} = 0$ at level α_1 , where α_1 is a tuning parameter. Let L^* denote the number of SNPs that passed this step.

3. Step 2: *Test for interactions*

Test for interactions at all pairs of L^* SNPs with Bonferroni correction for $\binom{L^*}{2}$ tests. Specifically, for a pair of SNPs (i, j) that passed Step 1(b), the interaction effect for these two SNPs in the k -th population is estimated by equation (1) within each study, and a test of overall interaction is

conducted based on $\hat{\beta}_3^{\text{overall}}$ defined in equation (2). The significance level is $\alpha / \binom{L^*}{2}$, where α is the nominal level for Type-I error.

Hereafter we refer the above procedure as KL-meta.

As acknowledged by its authors, KL can lead to non-monotonic power functions as the interaction effect β_3 increases (Kooperberg & LeBlanc, 2008). Similarly, non-monotonic power functions are to be expected for KL-meta. This is because when the parameters $(\beta_1^{(k)}, \beta_2^{(k)})$ and interaction effect $\beta_3^{(k)}$ in equation (1) are in opposite directions, the marginal effects $(\eta_1^{(k)}, \eta_2^{(k)})$ in equation (3) may be close to zero due to the cancellation between $(\beta_1^{(k)}, \beta_2^{(k)})$ and $\beta_3^{(k)}$. In such cases, the estimated overall marginal effects $\hat{\eta}_1^{\text{overall}}$ and $\hat{\eta}_2^{\text{overall}}$ will also be close to zero and, consequently, these two interacting SNPs will tend not to pass Step 1(b) even when an interaction effect is large.

A Novel Heterogeneity Based Test—The intuition for this procedure is that heterogeneity in marginal effects at a SNP across studies may be caused by interactions with other genetic loci that exhibit varying exposures (frequencies) among populations. Given sufficient population structure within or across studies, modeling this heterogeneity may contribute to discovery of interacting loci and provide a maximal utilization of information available from GWA studies. Our proposed procedure (YETI) is as follows:

1. Step 1(a): *Subgroup analysis, conducted within each study*

Estimation of the marginal effects is performed as in Step 1(a) of the KL-meta procedure.

2. Step 1(b): *Screening SNPs with significant heterogeneity among effect size estimates*

At the l -th SNP, we propose to conduct one of the following two tests for heterogeneity in marginal effects: $\eta_{1(l)}^{(1)} = \dots = \eta_{1(l)}^{(K)}$ (fixed-effect model), or $\text{var}(\eta_{1(l)}^{(k)})=0$ (random-effects model; $\eta_{1(l)}^{(k)}$ is treated as a random effect for $k = 1, \dots, K$) at level α_1 for $l = 1, \dots, L$. Such a test of heterogeneity can be conducted by using Wald test or likelihood ratio test. Let $L^\dagger$ denote the number of SNPs that passed this step.

3. Step 2: *Test for interactions*

Test the interactions between SNPs that passed Step 1(b) of YETI and *all* other SNPs with Bonferroni correction for $L^\dagger(L-1)$ tests within each study. Specifically, for a SNP that passed Step 1(b) of YETI, the interaction between this SNP and any distinct SNP in k -th population is quantified by equation (1), and a test of overall interaction is conducted based on $\hat{\beta}_3^{\text{overall}}$ defined in equation (2). The significance level is $\alpha / \{L^\dagger(L-1)\}$, where α is the nominal level for Type-I error.

The number of tests in step 2 of KL-meta, $\binom{L^*}{2}$, is typically smaller than the number of tests in step 2 of YETI $L^\dagger(L-1)$, especially when the number of SNPs L is relatively large. Since KL assumes both interacting loci exhibit significant marginal effects, it only needs to propagate to step 2 markers that passed step 1. However, for YETI, since an interaction may involve a locus that does not exhibit heterogeneity in effect size across studies, we need to test each marker that passes step 1 against all distinct markers in step 2. Interestingly, both the KL-meta and YETI methods can be shown to have the filtering property (Dai *et al.*, 2012) where the test statistics in step 1 can be used to filter out SNPs with low pooled marginal effects or a low degree of heterogeneity. For more details, please refer to Appendix 1 of the supplementary materials.

Computation

For a large number of SNPs, parallelization can reduce overall computation time. In fact, both KL-meta and YETI can be parallelized. We take YETI as an example. Suppose that one million SNPs are considered for possible interactions and there are K studies. Data from each study can be split into ten subsets with of approximately equal size, say

$$L_{S_1} = \{1, \dots, 10^5\}, L_{S_2} = \{10^5+1, \dots, 2 \times 10^5\}, \dots, L_{S_{10}} = \{9 \times 10^5+1, \dots, 10^6\}.$$

Step 1(a) and Step 1(b) of YETI can be conducted for each collection of SNPs L_{S_j} at the j -th computing node ($j = 1, \dots, 10$), at which tests can be carried out simultaneously. For each subset L_{S_j} , suppose there are $L_{S_j}^\dagger$ SNPs that passed Step 1(b) of YETI. The test of interaction

among these $L_{S_i}^\dagger$ SNPs with the other SNPs can be also carried out by multiple computing nodes simultaneously.

Working R code for YETI, along with a working example, can be obtained at <http://scheet.org/software.html>.

Results

Simulation Settings

We conduct simulation studies to evaluate the performance of KL-meta and YETI. We consider meta-analyses of two studies, each with sample sizes of 5000 cases and 5000 controls sampled from a homogeneous population. Although in practice the number of GWA studies available may exceed two, simulation studies with more populations (e.g., $K = 8$) result in similar relative comparisons among methods. Here we describe the meta-analysis of two studies for simplicity. We consider three scenarios depending on the size of main effects β_1 and β_2 , and the dependence between the “causal” SNPs. In the first scenario, the two causal SNPs are uncorrelated and the main effects are zero, i.e. $\beta_1^{(k)} = \beta_2^{(k)} = 0$ for $k = 1, 2$.

The interaction effects $\beta_3^{(1)}$ and $\beta_3^{(2)}$ in the two studies are set to be identical under all settings (hereafter referred to as β_3 for simplicity) and vary in the range of 0 to 1 for evaluation of the empirical Type-I error and power of the proposed procedures. As the interaction effects increase, the marginal effects $\eta_1^{(k)}$ and $\eta_2^{(k)}$ increase and are of the same direction (effect on phenotype). KL-meta should perform well in such a scenario because its identification of interaction relies on the detection of significant marginal effects for *both* SNPs in step 1. The second scenario is similar to the first one, except that the two causal SNPs are assumed to be negatively correlated with Pearson correlation coefficient of -0.2 . To obtain the negatively correlated risk alleles, we generate the two SNPs jointly from a multinomial distribution with $Pr(SNP1 = SNP2 = 1)$ calculated from marginal allele frequencies at SNP1 and SNP2 and the pre-specified correlation. Unlike the setting with independent SNPs, the marginal effects at the two SNPs induced by the interaction effect are different. For our example, one of the marginal effects is relatively large whilst the one marginal effect is close to zero. An empirical investigation of the marginal effects with corresponding interaction effects is summarized in Table S1 of the supplementary materials. It is expected that KL-meta may not perform as well because it relies heavily on the magnitude of the marginal effects of both causal SNPs, whereas the YETI procedure does not; however, YETI does of course rely on population differentiation. In the third scenario, we consider that the main effects of two causal SNPs are in the opposite direction of the interaction effect. Specifically, the main effects at two SNPs are equal and fixed at a negative value, i.e., $\beta_1^{(k)} = \beta_2^{(k)} = -0.25$ for $k = 1, 2$ in scenario 3(a), and $\beta_1^{(k)} = \beta_2^{(k)} = -0.45$ for $k = 1, 2$ in scenario 3(b). In these scenarios, the marginal effects are negative when the interaction effect is small, are close to 0 when the interaction effect is moderate, and then are positive again and increasing with an interaction effect beyond this moderate value. This illustrates the non-monotonic power function of KL-meta.

The allele frequencies at the two causal SNPs are identical *within* a study and are set at 0.1 for study 1 and 0.4 for study 2. The genotypes at the remaining non-causal SNPs are generated from a Bernoulli distribution with probability equal to one minus the square of the frequency of the major allele. Allele frequencies at non-causal SNPs are sampled from reference panels of The 1000 Genomes Project (here we used CEU and YRI) (1000 Genomes Project Consortium, 2012), with each panel providing the MAF distribution for one of the simulated GWA studies. To make the non risk and risk SNPs more similar, and thus mitigate any advantage, we excluded SNPs with MAFs below 0.1. We consider two “tuning” parameter values of $\alpha_1 = 0.002$ and 0.01, and the Type-I error is set at $\alpha = 0.05$. For each scenario, we consider settings with numbers of markers L being 50 or 1000. For each setting, 1000 replicates are conducted. In all scenarios above, genotypes are generated according to the described allele frequencies, independently at each SNP, with effects of linkage disequilibrium (LD) explored separately. Given the genotypes, a phenotype is generated from the logistic regression model as specified by equation (1), where the intercept β_0 is set at -2 for each study for a moderate prevalence (i.e., 11.9%) among noncarriers.

Simulation Results

Empirical Type-I errors of both KL-meta and YETI procedures are summarized in Table 1, with the step-1 filtering parameter α_1 set to 0.002 and 0.01 for $L = 50$ and 1000 SNPs. In practice, investigation with 50 SNPs is unrealistically small, but we include it to evaluate the impact of the number of markers. As expected, when a very small number of SNPs (i.e., 50) are analyzed, Type-I errors of both procedures are very conservative under all scenarios (except KL-meta in Scenarios 3(a) and 3(b)). This is because with a small number of SNPs, there are very few SNPs passing the filtering step 1, leading to conservative Type-I errors. In this case, the parameter $\alpha_1 = 0.002$ or 0.01 is still too stringent given the small number of SNPs, and the choice of tuning parameter has a large impact on the empirical Type-I errors. For Scenarios 3(a) and 3(b), although there is no interaction effect, the marginal effects (induced by the large main effects) are still substantial (approximately -0.3 for Scenarios 3(a), and -0.5 for Scenarios 3(b); see the first row of Tables S2 and S3). Thus, both causal SNPs typically pass step 1 of KL-meta procedure, leading to the Type-I errors close to the nominal level. Under the more practical setting of number of SNPs ($L = 1000$), the empirical Type-I errors are relatively close to the nominal level of 0.05, although the tuning parameter α_1 still has sizable impact on empirical Type-I errors. Nevertheless, both procedures with $\alpha_1 = 0.01$ control the Type-I errors satisfactorily. For larger numbers of SNPs ($L=2000$), the empirical Type-I error rates are close to the nominal level of 5% under all scenarios (simulation results not shown).

Power functions of both procedures for Scenario 1 are displayed in Figure 1, with increasing interaction effect β_3 . As expected, the KL-meta procedure has consistently higher power than the YETI procedure because the causal SNPs have increasingly larger marginal effects for *both* SNPs that are more likely to pass the filtering step. In addition, KL-meta tends to have fewer tests in step 2 (i.e., $L^* (L^* - 1)/2$) than the number of tests in the step 2 of YETI procedure (i.e., $L^\dagger (L - 1)$), since typically $L \gg L^*$ or $L^\dagger$. The tuning parameter α_1 seems to have a small impact on the power, especially when the number of SNPs L becomes large. As

the number of SNPs L increases, there is a sizable decrease in the power of both procedures. This can be explained by the more stringent correction for multiple tests. We note that the difference in power between KL-meta and YETI is caused both by different criteria of selecting SNPs in step 1, and also by a different searching strategy in step 2. To evaluate the impact of the searching strategy alone, we also conducted a comparison of KL-meta and YETI, each adapting the other test's searching strategy; see Figure S5 in the supplementary materials. Interestingly, we found that KL-meta has more power than its alternative searching strategy, and similarly for YETI. In other words, each method is optimal in their searching strategy at the 2nd stage (Step 2). This makes sense because their specific search strategies are corresponding to their underlying assumptions.

Power comparisons for both methods under Scenario 2, where two causal SNPs are negatively correlated (e.g. where the SNPs are nearby but the risk alleles on distinct haplotypes), are displayed in Figure 2. This could arise quite easily when the risk alleles are in LD. In this case, KL-meta is substantially underpowered. This is because one of the marginal effects of the two causal SNPs is close to zero, even when the interaction effects get larger. Thus, in this scenario, the YETI procedure is consistently more powerful than KL-meta. Another undesirable situation for KL-meta (not shown here) is that the effects at the same causal SNP are in opposite directions for two studies, which could be plausible when the haplotype tagging the risk variant in the two populations is at substantially different frequencies, i.e. a risk allele is on a common haplotype in one population but where that same haplotype is at lower frequency in another population. This situation results in small *overall* marginal effect due to cancellation between studies. In this case, the KL-meta procedure has limited power as well.

In Figure 3, we show the non-monotonicity of KL-meta under Scenarios 3(a) and 3(b). In the left panels, the marginal effects for both causal SNPs at both studies are zero when the interaction effect β_3 is 0.5, i.e., $\eta_1^{(k)} = \eta_2^{(k)} = 0$ for $k = 1, 2$ at $\beta_3 = 0.5$. At this value, SNPs will pass Step 1 of KL-meta procedure according to the Type-I error of the test, or the filtering parameter α_1 , and the procedure's power should not exceed the Type-I error. When the interaction effect deviates from 0.5, there exist negative or positive marginal effects, and as this deviation increases, there is an increase in power of the KL-meta procedure, except that the power is small when the interaction effect is close to 0 (i.e., approaching the Type-I error as $\beta_3 \approx 0$). This explains the “blip” or first mode of the multimodal power curve in the left column of Figure 3. The right panels of Figure 3 reveal similar findings with the marginal effects being zero at $\beta_3 = 1$. In this case, the power of the KL-meta procedure is close to 0 at $\beta_3 = 0$ and 1, leading to non-monotonic curves. (The right column of Figure 3 is showing the first mode only; at values of β_3 exceeding those shown, the power increases monotonically.) In contrast, the power curves of the YETI procedure are monotonic in all settings. This is because the YETI procedure is not based on the marginal effect sizes. The same phenomenon holds when β_1 and β_2 are of the same direction but opposite that of β_3 ; e.g., the protective effect of either SNP is reversed when both “risk” alleles are present. Since the YETI method relies on the heterogeneity in marginal effects across populations, it requires that the minor allele frequency (MAF) is different across populations. More specifically, due to the searching strategy of Step 2 of the YETI method, it only requires one of the two

interacting SNPs to have heterogeneous MAF across populations. For more details, please refer to the additional simulation results in the supplementary materials and Figure S5.

To investigate the impacts of the marginal effects for both procedures, we consider an additional simulation scenario where the main effect sizes are chosen such that the marginal effects of causal SNPs are fixed at 0.1 for all interaction effect sizes. The power curves are plotted in Figure S1 of the Supplementary Materials. In this case, the KL-meta procedure has low power due to the modest marginal effects. In contrast, the YETI procedure remains competitive, since the heterogeneity of allele frequencies is maintained. We also consider a similar setting as in Scenario 1 but the two causal SNPs are not directly genotyped. Instead, two surrogate SNPs that are in LD with disease-causal SNPs are genotyped. Figure S2 of the Supplementary Materials shows the power curves of both procedures. Similar findings are obtained in that the KL-meta procedure is still slightly more powerful than the YETI procedure. We observe a slight power loss for each procedure when comparing with a non-surrogate setting (i.e., Scenario 1), which is expected. Finally, we present the empirical investigation in power functions from both mega-analysis and meta-analysis via the KL procedure. As shown in Figure S3, both analyses have very similar power, which agrees with the theoretical results provided by Lin & Zeng (2010).

In summary, the KL-meta procedure has more power than the YETI procedure when the marginal effects at both causal SNPs are relatively large. On the other hand, the YETI procedure outperforms the KL-meta procedure in other situations such as when no more than one of causal SNPs has a large marginal effect, two causal SNPs are negatively correlated, or the main effects of causal SNPs are in opposite direction of the interaction effects. In addition, the YETI procedure, which is not based on the screening of marginal effects, always leads to monotonic power functions of interaction effects, whereas the KL-meta procedure can lead to poor or non-monotonic power functions. Thus, the YETI procedure is a useful complement to the KL-meta procedure.

Discussion

We have introduced two innovations by which the genetic epidemiologist may detect gene-gene interactions from GWA study data. The first was an extension of the method of Kooperberg & LeBlanc (2008), who proposed a constrained search for interactions among loci with significant marginal effects on phenotype. We adapted their procedure to accommodate GWA data from multiple studies, using a meta-analytic framework. The second innovation was to propose candidate loci for tests of interactions based on heterogeneity of effect size estimates, also using meta-analysis, with the logic being that other, interacting, loci will perturb the effect size estimates differentially if their allele frequencies differ among populations from which the study samples derived. That is, the interacting loci, if at differential allele frequencies (i.e., a type of exposure) among study populations, will exert differential influence on the effect size estimates of individual loci.

Here we have evaluated the relative performance of YETI in the setting of two GWA studies from differentiated populations. However, the procedure can handle an arbitrary number of populations. Aside from an increase in sample size from examining additional populations,

there exist at least two additional advantages of multiple populations. First, with more populations, there is a greater chance that one of the causal alleles will exhibit different allele frequencies between one of the populations and the others. Second, with more populations (e.g., 5), a random-effects meta-analysis may also be applied to reduce the degrees of freedom in the test of homogeneity; KL-meta would benefit from such a scenario, as well. We note that both KL-meta and YETI may be optimized by using marginal evidence for interaction from the first step to prioritize testing in the second step; one way to do this would be to partition the type-I error accordingly.

One concern about YETI could be a susceptibility to non-genetic confounders, i.e. environmental factors, that might differ among populations, especially since these are usually considered a nuisance in epidemiological studies of association. If known, these environmental factors can be adjusted through a meta-regression analysis in Step 1(b) of YETI procedure. If they are unknown or poorly characterized, while the first step could in fact be vulnerable to differential exposures from confounding factors among different studies, our 2-step procedure protects against this by testing for the proposed interactions *within* each study where the confounders would be uniform or for which knowledge may exist to properly adjust. It is worth noting that, since each study is examined separately in Step 2, although testing too many candidates would limit power, it would not inflate type-I error. In this vein, putative interacting candidate loci that pass Step 1 and fail to replicate within studies could be explained by the presence of additional genetic complexities, such as in higher-order genetic interactions or differential environmental exposures. In future work, a framework similar to ours could be adapted to specifically detect these influences.

Another possible concern could come from the very motivation to search for genetic interactions. While there have been documented examples, such as underlying clinical manifestation of Huntington's disease (Lee *et al.*, 2015), recent applications of linear mixed effect models have demonstrated that a large portion of estimated heritability can be explained by additive statistical models, thus indirectly calling into question the importance of epistatic interactions as a major source of "missing heritability", at least for some traits such as BMI and height (Yang *et al.*, 2015). Yet, these authors do not rule out epistasis and in fact one of the authors has contributed to demonstrating interactions affecting transcription (Hemani *et al.*, 2014); indeed, it has been argued that methods for epistasis should be employed for full assessment of heritability (Zuk *et al.*, 2012). Further, even when all marginal causal genetic factors have been identified, genetic interactions may reveal insights about disease architecture or underlying biology. Finally, it is not unclear if such broad conclusions of accounting for large proportions of heritability from methods such as GCTA are actually justified (Krishna Kumar *et al.*, 2016).

Since KL-meta selects SNPs based on pooled marginal effects, when the main effects exert effect on phenotype in the opposite direction of the interaction effect, there will be a particular interaction effect size that counteracts or neutralizes the single-marker effects on phenotype. At such combinations of parameters, the estimated marginal effects will be null and will not pass the first step of KL methods beyond the Type-I error rate; therefore, power is at its lowest point with a non-zero interaction effect, which is suboptimal and acknowledged by the authors of KL. In contrast, the power of YETI monotonically increases

in the magnitude of the interaction effect. In addition, since YETI screens the SNPs by the heterogeneity in marginal effects across studies, intuitively, KL-meta is preferred when populations in multiple studies are more homogeneous, whereas YETI gains favor when they are more heterogeneous.

In practice, it may require great differences among populations for the overall power of YETI to be comparable to methods that leverage mean effect sizes (rather than heterogeneity). Indeed, the motivation for YETI was not to replace these methods but to complement them, since combinations of parameters that lead to small marginal effect sizes will represent difficult scenarios for the detection of genetic interactions. Although naive methods that evaluate all possible pairs of tests will encounter the truly interacting loci, statistical significance will be difficult to obtain due to multiplicity. A limitation of YETI is that it mainly deals with this multiplicity by suggesting tests with more power. We do not necessarily expect interacting loci at differential frequencies among populations to be more likely to interact. That said, interactions involving small marginal effects should be subject to more limited negative selection, since selection will act on the average effects. Further, selection could act to induce allele frequency differences among populations if they are under differential environmental (i.e., selection) gradients. By contrast, the assumptions of KL are straightforward and intuitive. An illustration of this is that it requires some care to select advantageous scenarios for YETI compared to KL. It seems that scenarios most suitable to KL may in fact be the more common. Although that could be due to methods being designed to capture those, a detectable marginal effect is also a statistical consequence of a large interaction effect, unless the main effects have particular configurations. Still, several interesting scenarios are revealed, such as when the effects are correlated. Regardless, our goal here was to put forth a method that should help maximize the extraction of information from existing and expanding GWA studies, particularly of diverse samples, which may be especially relevant as more phenotypic and genetic data on samples worldwide continue to be collected, which will undoubtedly to be the case with globalization of existing technologies and increasing sophistication of electronic medical records.

It is encouraging that our two meta-analysis methods (KL-meta and YETI) are found to be complementary to each other. Specifically, when there is truly an interaction, the simulated data that were identified with interactions from these two procedures appear to be independent. This is not surprising, since the KL-meta procedure is based on the pooled mean while YETI is based on heterogeneity, and these two pieces of information are approximately orthogonal to each other. This is convenient, since our main objective was to characterize and harness the information contained in the heterogeneity across studies; YETI is meant to complement to the existing test KL or KL-meta. This relation between the tests suggests that combining two tests may offer a test that inherits the strength of both methods. We are pursuing this in concurrent work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to thank Chad Huff, Peng Wei, and two anonymous reviewers for helpful comments. This work was partially supported by NIH grants R01 HG 005859 (for PS, YL), R21 LM 012197 (for YC) and AHRQ grant R03 HS 022900 (for YC).

References

- 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. [PubMed: 23128226]
- Begum, Ferdouse; Ghosh, Debashis; Tseng, George C.; Feingold, Eleanor. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic acids research*. 2012; 40(9): 3777–3784. [PubMed: 22241776]
- Bureau, Alexandre; Dupuis, Josée; Falls, Kathleen; Lunetta, Kathryn L.; Hayward, Brooke; Keith, Tim P.; Van Eerdewegh, Paul. Identifying SNPs predictive of phenotype using random forests. *Genetic epidemiology*. 2005; 28(2):171–182. [PubMed: 15593090]
- Cantor, Rita M.; Lange, Kenneth; Sinsheimer, Janet S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *The American Journal of Human Genetics*. 2010; 86(1):6–22. [PubMed: 20074509]
- Chatterjee, Nilanjan; Kalaylioglu, Zeynep; Moslehi, Roxana; Peters, Ulrike; Wacholder, Sholom. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *The American Journal of Human Genetics*. 2006; 79(6):1002–1016. [PubMed: 17186459]
- Cordell, Heather J. Detecting gene-gene interactions that underlie human diseases. *Nature Reviews Genetics*. 2009; 10(6):392–404.
- Culverhouse, Robert; Suarez, Brian K.; Lin, Jennifer; Reich, Theodore. A perspective on epistasis: limits of models displaying no main effect. *The American Journal of Human Genetics*. 2002; 70(2): 461–471. [PubMed: 11791213]
- Dai, James Y.; Kooperberg, Charles; Leblanc, Michael; Prentice, Ross L. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika*. 2012; 99(4): 929–944. [PubMed: 23843674]
- DerSimonian, Rebecca; Laird, Nan. Meta-analysis in clinical trials. *Controlled clinical trials*. 1986; 7(3):177–188. [PubMed: 3802833]
- Duerr, Richard H.; Taylor, Kent D.; Brant, Steven R.; Rioux, John D.; Silverberg, Mark S.; Daly, Mark J.; Steinhardt, A Hillary; Abraham, Clara; Regueiro, Miguel; Griffiths, Anne, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *science*. 2006; 314(5804):1461–1463. [PubMed: 17068223]
- Eichler, Evan E.; Flint, Jonathan; Gibson, Greg; Kong, Augustine; Leal, Suzanne M.; Moore, Jason H.; Nadeau, Joseph H. Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics*. 2010; 11(6):446–450.
- Evangelou, Evangelos; Ioannidis, John PA. Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*. 2013; 14(6):379–389.
- Evans, David M.; Marchini, Jonathan; Morris, Andrew P.; Cardon, Lon R. Two-stage two-locus models in genome-wide association. *PLoS Genetics*. 2006; 2(9):e157. [PubMed: 17002500]
- Guerra, Rudy; Goldstein, Darlene R. Meta-analysis and combining information in genetics and genomics. CRC Press; 2009.
- Hemani G, Shakhbazov K, Westra HJ, Esko T, Henders AK, McRae AF, Yang J, Gibson G, Martin NG, Metspalu A, Franke L, Montgomery GW, Visscher PM, Powell JE. Detection and replication of epistasis influencing transcription in humans. *Nature*. 2014; 508(7495):249–253. [PubMed: 24572353]
- Jiang, Rui; Tang, Wanwan; Wu, Xuebing; Fu, Wenhui. A random forest approach to the detection of epistatic interactions in case-control studies. *BMC bioinformatics*. 2009; 10(Suppl 1):S65. [PubMed: 19208169]

- Kononenko, Igor. Pages 171–182 of: Machine Learning: ECML-94. Springer; 1994. Estimating attributes: analysis and extensions of RELIEF.
- Kooperberg, Charles; LeBlanc, Michael. Increasing the power of identifying gene× gene interactions in genome-wide association studies. *Genetic epidemiology*. 2008; 32(3):255–263. [PubMed: 18200600]
- Krishna Kumar S, Feldman MW, Rehkopf DH, Tuljapurkar S. Limitations of GCTA as a solution to the missing heritability problem. *Proc Natl Acad Sci USA*. 2016; 113(1):61–70.
- Lee JM, Wheeler VC, Chao MJ, Vonsattel JP, Pinto RM, Lucente D, Abu-Elneel K, Ramos EM, Mysore JS, Gillis T, MacDonald ME, Gusella JF, Harold D, Stone TC, Escott-Price V, Han J, Vedernikov A, Holmans P, Jones L, Kwak S, Mahmoudi M, Orth M, Landwehrmeyer GB, Paulsen JS, Dorsey ER, Shoulson I, Myers RH. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell*. 2015; 162(3):516–526. [PubMed: 26232222]
- Lin DY, Zeng D. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika*. 2010; 97(2):321–332. [PubMed: 23049122]
- Lunetta, Kathryn L.; Hayward, L Brooke; Segal, Jonathan; Van Eerdewegh, Paul. Screening large-scale association study data: exploiting interactions using random forests. *BMC genetics*. 2004; 5(1):32. [PubMed: 15588316]
- Manolio, Teri A.; Collins, Francis S.; Cox, Nancy J.; Goldstein, David B.; Hindorff, Lucia A.; Hunter, David J.; McCarthy, Mark I.; Ramos, Erin M.; Cardon, Lon R.; Chakravarti, Aravinda, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; 461(7265):747–753. [PubMed: 19812666]
- McCarthy, Mark I.; Abecasis, Gonçalo R.; Cardon, Lon R.; Goldstein, David B.; Little, Julian; Ioannidis, John PA.; Hirschhorn, Joel N. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*. 2008; 9(5):356–369.
- Moore, Jason H.; White, Bill C. Pages 166–175 of: Evolutionary computation, machine learning and data mining in bioinformatics. Springer; 2007. Tuning ReliefF for genome-wide genetic analysis.
- Moore, Jason H.; Williams, Scott M. Epistasis and its implications for personal genetics. *The American Journal of Human Genetics*. 2009; 85(3):309–320. [PubMed: 19733727]
- Murcray, Cassandra E.; Lewinger, Juan Pablo; Gauderman, W James. Gene-environment interaction in genome-wide association studies. *American journal of epidemiology*. 2009; 169(2):219–226. [PubMed: 19022827]
- Park, Mee Young; Hastie, Trevor. Penalized logistic regression for detecting gene interactions. *Biostatistics*. 2008; 9(1):30–50. [PubMed: 17429103]
- Ritchie, Marylyn D.; Hahn, Lance W.; Roodi, Nady; Bailey, L Renee; Dupont, William D.; Parl, Fritz F.; Moore, Jason H. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *The American Journal of Human Genetics*. 2001; 69(1):138–147. [PubMed: 11404819]
- Ritchie, Marylyn D.; White, Bill C.; Parker, Joel S.; Hahn, Lance W.; Moore, Jason H. Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases. *BMC bioinformatics*. 2003; 4(1):28. [PubMed: 12846935]
- Saxena, Richa; Voight, Benjamin F.; Lyssenko, Valeriya; Burt, Noël P.; de Bakker, Paul IW.; Chen, Hong; Roix, Jeffrey J.; Kathiresan, Sekar; Hirschhorn, Joel N.; Daly, Mark J., et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007; 316(5829):1331–1336. [PubMed: 17463246]
- Sladek, Robert; Rocheleau, Ghislain; Rung, Johan; Dina, Christian; Shen, Lishuang; Serre, David; Boutin, Philippe; Vincent, Daniel; Belisle, Alexandre; Hadjadj, Samy, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007; 445(7130):881–885. [PubMed: 17293876]
- Song, Chi; Tseng, George C. Hypothesis setting and order statistic for robust genomic meta-analysis. *The annals of applied statistics*. 2014; 8(2):777–800. [PubMed: 25383132]
- Tanck, Michael WT.; Wouter Jukema, J.; Zwinderman, Aeilko H. Simultaneous estimation of gene-gene and gene-environment interactions for numerous loci using double penalized log-likelihood. *Genetic epidemiology*. 2006; 30(8):645–651. [PubMed: 16917921]

- Van Steen, Kristel. Travelling the world of gene–gene interactions. *Briefings in Bioinformatics*. 2011; 13(1):1–19. [PubMed: 21441561]
- Wei, Wen-Hua; Hemani, Gibran; Haley, Chris S. Detecting epistasis in human complex traits. *Nature Reviews Genetics*. 2014; 15:722–733.
- Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, Robinson MR, Perry JR, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, Esko T, Milani L, Magi R, Metspalu A, Hamsten A, Magnusson PK, Pedersen NL, Ingelsson E, Soranzo N, Keller MC, Wray NR, Goddard ME, Visscher PM. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet*. 2015; 47(10):1114–1120. [PubMed: 26323059]
- Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci USA*. 2012; 109(4):1193–1198. [PubMed: 22223662]

Figure 1.

Power curves for identifying an interaction when $\beta_1^{(k)} = \beta_2^{(k)} = 0$ for $k = 1, 2$, for various interaction effects between two uncorrelated causal SNPs.

Figure 2.

Power curves for identifying an interaction when $\beta_1^{(k)} = \beta_2^{(k)} = 0$ for $k = 1, 2$, for various interaction effects between two negatively correlated causal SNPs.

Figure 3.

Power curves for identifying an interaction when $\beta_1^{(k)} = \beta_2^{(k)} = -0.25$ for $k = 1, 2$ in (a) and $\beta_1^{(k)} = \beta_2^{(k)} = -0.45$ for $k = 1, 2$ in (b), for various interaction effects between two uncorrelated causal SNPs.

Empirical Type-I error rates (%) of the KL-meta and YETI procedures with the nominal level of 5% based on 1000 simulations. Scenario 1: $\beta_1^{(k)} = \beta_2^{(k)} = 0$; Scenario 2: $\beta_1^{(k)} = \beta_2^{(k)} = 0$ and two causal SNPs have a correlation of -0.2 ; Scenario 3: (a) $\beta_1^{(k)} = \beta_2^{(k)} = -0.25$, and (b) $\beta_1^{(k)} = \beta_2^{(k)} = -0.45$, for $k = 1, 2$.

Table 1

Scenario 1			Scenario 2			Scenario 3(a)			Scenario 3(b)		
L	α_1	KL-meta	YETI	KL-meta	YETI	KL-meta	YETI	KL-meta	YETI	KL-meta	YETI
50	0.002	0.1	0.1	0.2	0.4	5.1	0.4	5.2	0.6		
	0.01	0.3	0.3	1.7	1.8	5.2	1.7	5.4	2.3		
1000	0.002	2.9	5.2	2.5	3.9	4.4	4.7	5.7	4.2		
	0.01	4.9	4.7	5.2	6.2	4.9	5.8	4.5	5.5		