Subset-Based Analysis using Gene-Environment Interactions for Discovery of Genetic Associations across Multiple Studies or Phenotypes

Youfei Yu[1,¶], **Lu Xia**[1,¶], **Seunggeun Lee**[1,2], **Xiang Zhou**[1,2], **Heather M Stringham**[1,2], **Michael Boehnke**[1,2], **Bhramar Mukherjee**[1,2,*]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, United States of America

[2]Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, United States of America

Abstract

Objectives: Classical methods for combining summary data from genome-wide association studies (GWAS) only use marginal genetic effects and power can be compromised in the presence of heterogeneity. We aim to enhance the discovery of novel associated loci in the presence of heterogeneity of genetic effects in sub-groups defined by an environmental factor.

Methods: We present a p-value Assisted Subset Testing for Associations (pASTA) framework that generalizes the previously proposed *as*sociation analysis based on sub*set*s (ASSET) method by incorporating gene-environment (G-E) interactions into the testing procedure. We conduct simulation studies and provide two data examples.

Results: Simulation studies show that our proposal is more powerful than methods based on marginal associations in the presence of G-E interactions and maintains comparable power even in their absence. Both data examples demonstrate that our method can increase power to detect overall genetic associations and identify novel studies/phenotypes that contribute to the association.

Conclusions: Our proposed method can be a useful screening tool to identify candidate single nucleotide polymorphisms (SNPs) that are potentially associated with the trait(s) of interest for further validation. It also allows researchers to determine the most probable subset of traits that exhibit genetic associations in addition to the enhancement of power.

[*]Correspondence: bhramar@umich.edu, Address: Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, USA. Phone: (734) 764-6544, Fax: (734) 763-2215.
[¶]These authors contributed equally to this work.

The authors declare no conflicts of interest.

Web Resources
Current annotated code, https://github.com/youfeiyu/pASTA/

INTRODUCTION

Genome-wide association studies (GWAS) are popular epidemiologic tools for studying the genetic architecture underlying a phenotypic trait [1]. Meta-analysis is a commonly used approach to combine genetic associations across multiple independent studies [2]. Fixed-effect meta-analysis based on summary statistics is known to retain full efficiency of an analysis based on individual level data [3]. Since fixed-effect meta-analysis assumes common effect sizes across studies, the power of the fixed-effect method may be compromised when there is between-study heterogeneity in the effect size of the association. Han and Eskin [4] have proposed a random-effect model with a likelihood ratio test to account for such heterogeneity. More recently, methods for aggregating association signals across multiple related phenotypes have also become more popular [5,6].

The *as*sociation analysis based on sub*set*s (ASSET) test, proposed by Bhattacharjee et al. [7], uses summary statistics from individual association analysis to develop a powerful test that allows the existence of both a subset of null results and effects in opposite directions in different individual tests. Unlike the random-effect model proposed by Han and Eskin [4], ASSET assumes fixed effects for individual associations, but still allows for heterogeneity in the data. This approach essentially explores all possible non-empty subsets of available studies/traits and searches for the one that yields the strongest evidence of association while adjusting for multiple testing during such exhaustive search. One appealing feature of ASSET is that the most probable subset of studies/traits that exhibit genetic association can be identified in addition to the enhancement of power of detecting association signals. This method has been widely used in analyses of associations across multiple studies and phenotypes [8,9].

Exploiting gene-environment (G-E) interaction may help identify genetic variants that do not demonstrate very strong marginal effects due to environmental heterogeneity but may have stronger genetic effects in sub-groups defined by certain levels of an exposure [10–13]. Like most commonly used methods that screen only for marginal associations, ASSET does not account for potential G-E interactions in the testing procedure. If a genetic variant only affects a subgroup, for example the exposed group, then in the presence of such pure G-E interaction, the power of ASSET can be largely compromised. Several existing versions of 2-degree-of-freedom (2-df) tests for discovering genetic associations take G-E interactions into account. Kraft et al. [12] proposed a likelihood ratio test that jointly tests for genetic main effects and G-E interactions in case-control studies. Dai et al. [11] proposed a slightly different approach that is based on the sum of two Wald chi-square statistics for detecting marginal effects combined with G-E interactions. The latter approach has the advantage of incorporating G-E independence into the test statistic for G-E interaction in case-control studies to achieve increased power [14,15]. The literature on meta-analysis of G-E interactions or joint tests remains relatively limited [16–18], and none of these approaches attempt to specifically identify a subset of studies/traits that are most likely to have genetic associations and/or interactions. Current literature shows that the 2-df tests mentioned above can offer enhanced power in the presence of G-E interactions and do not lead to significant loss of power even when the interactions are absent [16,18].

In this paper, we propose a modification to ASSET that incorporates G-E interactions to increase the power of detecting genetic associations in combining summary data across studies of a given phenotype (say type 2 diabetes) or across multiple related phenotypes (say a set of lipid traits) measured in different studies. Our approach is similar to ASSET in terms of searching over all possible combinations of studies/phenotypes and thus inherits its advantage of identifying the maximally associated subset along with increased power for detecting the overall associations. The proposed framework uses p-values from the underlying tests as input instead of Z-statistics that are used by ASSET, and is therefore referred to as p-value Assisted Subset Testing for Associations (pASTA). To be able to incorporate G-E interactions, pASTA uses 2-df tests as the underlying test statistics to be combined. In fact, the p-values can result from any statistical tests, including tests with multiple degrees of freedom. On the other hand, by using p-values, one loses the ability to incorporate the signs of coefficients (for both genetic main effect and G-E interaction) and the method is intrinsically two-sided. Since the key input variables are the p-values from underlying tests, pASTA can be used under any commonly used epidemiologic study design and we present our simulation studies and data examples for both case-control and cohort studies. Like ASSET, the analytical method in pASTA can handle a set of correlated p-values, enabling its validity in studies with overlapping subjects or multiple phenotypes measured on the same set of subjects.

The rest of the paper is organized as follows. We first describe the construction of our test statistic and derivation of the associated p-value. We present extensive simulation studies to compare the performance of our proposed method to ASSET and Fisher's combined p-value approaches [19] with and without G-E interactions under various parameter settings. Specifically, we examine the type I error rates and the power to detect (1) the truly associated variant and (2) the truly associated subset of studies/traits for the given variant. We illustrate our proposed method with two data applications. The first is a meta-analysis of six case-control studies of type 2 diabetes (T2D) among European population (three are Finnish, one is German, and two are Norwegian), with 4,422 cases and 5,202 controls. We focus on two single nucleotide polymorphisms (SNPs) in the *FTO* gene (MIM 610966) related to obesity and their interactions with body mass index (BMI). The second application is a multiple-phenotype analysis of nine lipid-related quantitative traits using data of 5,123 individuals from the North Finland Birth Cohort 1966 (NFBC1966) Study. We investigate the association between two SNPs discovered by GWAS, one near the *LPL* gene (MIM 609708) and the other in the *APOB* gene (MIM 107730), and the nine lipid traits while taking the SNP-BMI interactions into consideration. We conclude the paper with a discussion.

MATERIAL AND METHODS

Notation

We consider K studies, each with sample size n_k. We allow the studies to have a set of overlapping subjects. Let Y_{ki}, G_{ki}, and E_{ki} denote the underlying phenotype, the given genetic variant and the environmental factor, respectively, measured for the ith subject in the kth study. The trait(s) Y can either be disease status in a case-control study or binary or

quantitative trait(s) measured in a cohort study. The genetic variant G may be coded as binary (under a dominant or recessive susceptibility model) or as allele count (under the additive model), and for the latter case, we treat G as continuous dosage. The environmental factor E can either be categorical or continuous. We mainly consider two models for each of the K studies: the marginal genetic association model

$$g[E(Y_{ki}|G_{ki})] = \alpha_0^{(k)} + \alpha_G^{(k)} \cdot G_{ki} \qquad \text{(Model 1)}$$

and the joint model with G-E interaction

$$g[E(Y_{ki}|G_{ki}, E_{ki})] = \beta_0^{(k)} + \beta_G^{(k)} \cdot G_{ki} + \beta_E^{(k)} \cdot E_{ki} + \beta_{GE}^{(k)} \cdot G_{ki} \cdot E_{ki} \qquad \text{(Model 2)}$$

where $i = 1, \cdots, n_k$ and $k = 1, \cdots, K$. The choice of the link function $g(\cdot)$ depends upon the type of Y. Adjusted covariates are dropped from the presentation for simplicity of notations. Note that the K studies in this setup can be replaced by K traits in a single study of size n, and the inferential procedure for each of the K traits stays the same with $n_k = n$ for each $k = 1, \cdots, K$.

We primarily consider three types of association tests for each of the K studies: 1) marginal genetic effect in Model 1 (MA), with null hypothesis (H_{0k}) being $\alpha_G^{(k)} = 0$; 2) genetic main effect and G-E interaction in Model 2 (JOINT) [12], with H_{0k} being $\beta_G^{(k)} = \beta_{GE}^{(k)} = 0$; and 3) marginal genetic effect in Model 1 and G-E interaction in Model 2 (MA+GE) [11], with H_{0k} being $\alpha_G^{(k)} = \beta_{GE}^{(k)} = 0$. MA is a 1-df test while JOINT and MA+GE are 2-df tests. A summary description of the different tests and their corresponding null hypotheses, test statistics and distributions under the null are presented in Appendix Table A1. The test of the form MA +GE is typically the sum of two Wald chi-square test statistics that are known to be asymptotically independent and can thus be combined to yield a 2-df chi-square statistic [20].

For a case-control study, the coefficient of interaction $\beta_{GE}^{(k)}$ can be estimated using case-control (CC), case-only (CO) [14], or empirical Bayes (EB) [15] estimators, respectively denoted as $\hat{\beta}_{CC}^{(k)}$, $\hat{\beta}_{CO}^{(k)}$ and $\hat{\beta}_{EB}^{(k)}$. The CO estimator increases the power of detecting G-E interaction by using G-E independence but incurs inflated type I error under violations of such independence The EB estimator can be viewed as a linear combination of CC and CO estimators, and it improves power to detect G-E interaction when compared to standard CC analysis and provides better control of type I error when compared to a CO analysis. It has been proved that $\hat{\beta}_{CC}^{(k)}$ and $\hat{\alpha}_G^{(k)}$ are asymptotically independent, and so are $\hat{\beta}_{EB}^{(k)}$ and $\hat{\alpha}_G^{(k)}$ when we assume G-E independence and a rare disease [20]. As such, for case-control studies we use $\hat{\beta}_{CC}^{(k)}$ and $\hat{\beta}_{EB}^{(k)}$ separately as two possible choices for the estimator of $\beta_{GE}^{(k)}$ when computing the test statistic of MA+GE, and denote these two types of tests as MA+CC and MA+EB, respectively [11]. Note that the use of G-E independence requires the use of a retrospective likelihood of the form $(G, E)|D$, which is not valid unless the sampling is

conditional on D. A cohort analysis is always based on $D|(G, E)$ and we cannot model the stochastic distribution of (G, E). Therefore, for a cohort study with multiple quantitative or binary phenotypes, the use of G-E independence does not lead to enhanced power and as such the EB estimator does not apply.

p-value Assisted Subset Testing for Associations (pASTA)

Regardless of the test statistic chosen for each study, the underlying analytical framework of pASTA starts with a set of p-values: p_k, $k = 1, \cdots, K$. In our setting they are obtained from the three tests, MA, JOINT, or MA+GE. For pASTA, we define the Z-statistic for the kth study as

$$Z_k = -\Phi^{-1}(p_k),$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) corresponding to the standard normal distribution. That is, smaller p-values indicate larger Z-statistics, and the null hypothesis H_{0k} should be rejected if Z_k is large in the positive direction. For any non-empty subset $S \subseteq \{1, \cdots, K\}$, let

$$Z(S) = \sum_{k \in S} \sqrt{\pi_k(S)} Z_k,$$

where we consider $\sqrt{\pi_k(S)} = \sqrt{n_k / \sum_{k \in S} n_k}$ as the weight of Z_k. Other options for the weights $\sqrt{\pi_k(S)}$ can also be accommodated in this framework. For example, when cases and controls are not of equal numbers, n_k can be substituted by the effective sample size [2]. The test statistic for evaluating the overall association of a given genetic variant is defined as

$$Z_{meta} = \max_{S \in \mathcal{S}} Z(S),$$

and is obtained by searching over all possible $|\mathcal{S}| = 2^K - 1$ non-empty subsets of studies.

It is challenging to characterize the exact distribution of Z_{meta} in a closed form, especially when the number of studies K is large. Therefore, we borrow the idea followed in ASSET and use the discrete local maxima (DLM) procedure [21] to obtain a conservative upper bound for the meta-analytic combined p-value. The DLM procedure automatically accounts for the multiple comparison issue. The detailed derivation of the combined p-value based on Z_{meta} is included in Appendix. In brief, let $\gamma \in \Gamma$ be a K-dimensional vector, each of whose components takes value in $\{0, 1\}$, and let S_γ be the subset of studies whose corresponding coordinates in γ are 1. A neighbor of S_γ is defined as a subset in $\mathcal{S}$ that can be obtained by adding or dropping one study to or from S_γ. Then $Z(S_{\gamma*})$ is called a local maxima if it is greater than all $Z(S_\gamma)$'s where S_γ refers to any possible neighbor of $S_{\gamma*}$. For an observed test statistic $Z_{meta} = T_{obs}$, the DLM approximated p-value can be expressed as (see Equation A1 in Appendix)

$$p_{DLM} = \int_{T_{obs}}^{\infty} \prod_{k=1}^{K} P\big(Z(S_\gamma \pm k) < z \,\big|\, Z(S_\gamma) = z\big) h(z) dz,$$

where $h(\cdot)$ is the probability distribution function of $Z(S_\gamma)$. This DLM-based p-value is actually an upper bound to the exact one (Equation A1 in Appendix). Hence the overall procedure is conservative in terms of type I error.

Each of the terms of the form $P(Z(S_\gamma \pm k) < z \,|\, Z(S_\gamma) = z)$ can be evaluated based on the conditional normal distribution of Z_k given $Z(S_\gamma)$ (Equations A2 and A3 in Appendix). Thus all we need to evaluate is the joint bivariate distributions of Z_k and $Z(S_\gamma)$ for $k = 1, \cdots, K$. Note that Z_k follows the standard normal distribution under the null hypothesis that there is no association (and interaction) in study k. The set of Z_k's are independent if the studies are independent. For multiple studies with overlapping subjects or multiple phenotypes measured on the same set of individuals, Z_k's will be correlated with variance one and correlation matrix Σ with entries say, $\rho_{kk'} = Corr(Z_k, Z_{k'})$.

Let $\boldsymbol{L} = (L_1, \cdots, L_K)$, where $L_i = 1$ if the i-th study is in S_γ and $L_i = 0$ otherwise. Let $\boldsymbol{Z} = (Z_1, \cdots, Z_K)$ follow a multivariate normal distribution $MVN_K(\boldsymbol{0}, \Sigma)$. In particular, if all studies/traits are independent, then Σ is a $K \times K$ identity matrix. Let $\boldsymbol{w} = (w_1, \cdots, w_K)$ be the vector of weights for the element in $\boldsymbol{L}$. $w_i = \sqrt{\dfrac{n_i}{\sum_{j \in S_\gamma} n_j}}$ if $L_i = 1$ and 0 if $L_i = 0$. Let $\boldsymbol{e}_k$ be a vector of length K where the kth element equals 1 and 0 elsewhere, e.g. $\boldsymbol{e_2} = (0,1,0,\cdots,0)^T$. Let $A_k = \begin{pmatrix} e_k^T \\ \omega^T \end{pmatrix}$, which is a $2 \times K$ matrix. Under the null hypothesis,

$$\begin{pmatrix} Z_k \\ Z(S_\gamma) \end{pmatrix} = A_k Z \sim MVN_2\Big(0, A_k \sum A_k^T\Big).$$

Several methods have been proposed to estimate the correlation matrix Σ. Lin and Sullivan ([22] provided a formula to estimate the correlation between a pair of studies using the information of shared cases and controls when the number of overlapping subjects is known. They also discussed a modified version of the formula for quantitative traits in a cross-sectional study. Bhattacharjee et al. [7] extended Lin and Sullivan's formula for case-control studies to accommodate the situation where cases of one study serves as controls of another. However, these estimates will likely be inaccurate when the overlap fraction is large, and therefore do not directly apply to multiple-phenotype analysis, where the phenotypes are measured on the same set of subjects and the overlap fraction is 100%. In this case, since large positive correlation between phenotypic measurements usually indicates large positive correlation between Z-statistics, an intuitive way to approximate the correlations of Z_k's is to use the phenotypic correlations [23], as is done in ASSET. We adopt this approach for our simulation studies and the second data example to estimate the correlation matrix Σ corresponding to $\boldsymbol{Z}$. In other words, if $\widehat{corr}(Y_1, \cdots, Y_k) = S$, then we let $\hat{\Sigma} = S$.

Comparison of pASTA and ASSET

We would like to point out a few features of pASTA as compared to ASSET when one is only interested in marginal genetic association estimated by Model 1. First, ASSET and pASTA use the *Z*-statistics in different ways (Appendix Table A2). The *Z*-statistics used in pASTA no longer differentiate between possible directions of association, while the signs of the *Z*-statistics used in ASSET are consistent with those of $\hat{\alpha}_G^{(k)}$'s. ASSET can detect genetic associations in either one direction (one-sided) or both directions (two-sided). For the latter case, ASSET searches for the subsets of studies that yield the strongest evidence separately in both directions. It obtains a p-value for each direction, and combines the two p-values using Fisher's combined p-value method [19]. pASTA does not have this feature. A more detailed comparison of the *Z*-statistics, test statistics and analytic expression for the DLM-based p-values of the two methods are presented in Appendix Table A2. The extension in this paper is mainly to incorporate interaction while testing association using 2-df tests in the ASSET framework via the use of p-values.

Simulation studies

Simulation Study 1: Combining Signals Across Multiple Independent Case-Control Studies

Methods Considered.: We first assess the performance of the proposed pASTA approach for combining summary results from multiple independent case-control studies that have no overlapping subjects or relatedness. We compare pASTA to three classes of alternatives (Table 1): ASSET, Fisher's combined p-value method [19], and the gold standard test which assumes that the true subset of non-null studies and the true model from which the data are generated are known a priori, to reflect the maximum achievable power for benchmarking each method. Given K independent p-values $p_1, \cdots, p_K$, the chi-square test statistic of Fisher's method follows a $\chi^2_{df = 2K}$ distribution under the global null. Bhattacharjee et al. [7] has already compared inverse-variance weighted meta-analysis with ASSET and Fisher's method, showing that the former is not as powerful as the latter two. Therefore, we do not include standard fixed effect inverse-variance weighted meta-analysis method in our comparative study unless otherwise noted. We use several choices of 2-df tests as described in Table 2 and expanded in Appendix Table A1.

Data Generative Model.: We consider $K = 10$ independent studies with 6,000 cases and 6,000 controls. The disease status D, the genetic marker G and the environmental factor E are assumed binary (1 for presence and 0 for absence) for simplicity. Following the standard approach of simulating case-control data with G-E interaction [10], we generate the data from a logistic regression model

$$\log\frac{P(D_{ki} = 1 | G_{ki}, E_{ki})}{P(D_{ki} = 0 | G_{ki}, E_{ki})} = \beta_0^{(k)} + \beta_G^{(k)} \cdot G_{ki} + \beta_E^{(k)} \cdot E_{ki} + \beta_{GE}^{(k)} \cdot G_{ki} \cdot E_{ki}$$

for $k = 1, \cdots, K$. In addition, we let θ_{GE}, p_G and p_E denote the G-E association, the frequency of $G = 1$ (this probability depends on the frequency of the risk allele, we only consider a dominant model in the simulation study) and the prevalence of environmental factor ($E = 1$)

in controls, respectively. We specify the parameters θ_{GE}, p_G, p_E, $\beta_G^{(k)}$, $\beta_E^{(k)}$, and $\beta_{GE}^{(k)}$, and determine the odds ratios of marginal genetic association (OR_G) and environmental association (OR_E) using the formula in Boonstra et al. [10]. In all of our simulation studies, we assume G-E independence among the control population, i.e., $\theta_{GE} = 0$.

Evaluation of Type I Error.: To assess the type I error rate of each method in the presence or absence of environmental effects, we generate data under the null hypothesis of no genetic effects, namely, set $\beta_G^{(k)} = \beta_{GE}^{(k)} = 0$ in Model 2. We consider two scenarios (Appendix Table A3), one with no environmental factor and hence no G-E interaction and the other with $OR_E = 1.4$. All ten studies share the same parameters in each scenario (Appendix Table A3). We evaluate the type I errors at three levels of significance: $\alpha = 10^{-2}$ and 10^{-3} resembling a candidate gene study, and 10^{-5} resembling large scale exploration. We simulate 5×10^5 replicated datasets for each scenario.

Metrics for Power Analysis.: We compare the power of pASTA with the other methods displayed in Table 1 from two perspectives: the power of detecting overall genetic associations and the accuracy of identifying the exact subset of non-null studies. Here the non-null studies refer to studies in which either $\alpha_G^{(k)} \neq 0$ (and hence $\beta_G^{(k)} \neq 0$) or $\beta_{GE}^{(k)} \neq 0$. The specific definitions of the two metrics are given in Appendix. We also report the sensitivity and specificity of identifying the subset of non-null studies, defined as the proportion of non-null studies that are correctly identified by subset-based approaches, and the proportion of truly null studies that are declared null, respectively.

Power for Detecting Marginal Association.: Prior to consideration of G-E interaction we extensively investigate the relative performance of ASSET and pASTA(MA) in detecting marginal genetic associations in a case-control setting where the data are generated under Model 1. We consider two scenarios in this simulation, each with $K = 10$ independent studies and $k_0 = 4$ of them having true associations. The prevalence of the genetic variant p_G is 0.2. In the first scenario, we assume that the marginal genetic effects in the four non-null studies are all in the positive direction. In the second, we reverse the sign of the log-odds ratio parameter of genetic effect (β_G) in two of the four studies. In both scenarios, the absolute values of log odds ratio of marginal associations $\left| \alpha_G^{(k)} \right|$ are the same for all non-null studies, ranging from log 1.05 to log 1.2, and $OR_G^{(k)}$ is incremented by 0.01. Since in reality it is possible that all studies have non-null signals, whether constant or varying across studies, we also carry out a simulation study where $k_0 = 10$. For this particular simulation setting, we include fixed-effect meta-analysis method in addition to the 1-df tests described in Table 1. For all the settings described above, we use $\alpha = 0.001$ as the significance level, which resembles a candidate gene study with 50 SNPs, and simulate 2,000 datasets for each parameter configuration.

Power for Detecting Association After Incorporating Interaction.: We then consider both G and E in our model and the whole set of 1-df and 2-df tests as listed in Table 1. We consider five scenarios (Appendix Table A4) to evaluate the power of pASTA. Scenario 1

assumes no G-E interaction and a moderate marginal genetic effect (OR_G). In Scenario 2, we set parameters to induce large G-E interaction but small marginal genetic effect. Scenario 3 considers the same magnitude of interaction as Scenario 2, but with zero genetic and environmental main effects, which result in a smaller marginal genetic effect compared to Scenario 2. Scenario 4 assumes strong G-E interaction and protective genetic main effect, the two of which cancel out and result in an odds ratio of marginal genetic effect close to one. Scenario 5 considers negative interaction and positive genetic main effect, which lead to a nearly-null marginal genetic effect. The last two scenarios represent less common case-control settings, where the genetic associations exist in opposite directions for exposed and unexposed groups. The number of non-null studies ranges from two to seven. Simulations are repeated 2,000 times for each scenario, and significance level is $\alpha = 0.001$.

Simulation Study 2: A Multiple-Phenotype Analysis in a Cohort Study

Methods Considered.: We compare the performance of pASTA and ASSET in a cohort study with a set of multiple correlated traits. In this case, the test statistics of Fisher's method no longer follows a $\chi^2_{df = 2K}$ distribution under the null hypotheses due to the correlations among $p_1, \cdots, p_K$. Resampling-based approaches can be used to obtain the null distribution when correlations exist [24], but in this situation we are dealing with significance level at the magnitude of 10^{-5} or smaller. It is computationally intensive to obtain a null distribution to evaluate the p-value at such a stringent threshold. Therefore, we exclude comparison with Fisher's method from Simulation Study 2.

Data Generative Model.: We consider a multiple-phenotype study with 6,000 subjects and $K = 9$ correlated traits. The outcome Y is generated from a multivariate linear regression model

$$Y_{ki} = \gamma_0^{(k)} + \gamma_G^{(k)} G_{ki} + \gamma_E^{(k)} E_{ki} + \gamma_{GE}^{(k)} G_{ki} \cdot E_{ki} + \epsilon_{ki},$$

where $(\epsilon_{1i}, \cdots, \epsilon_{Ki})^T \sim MVN(0, \Omega)$. We use the empirical phenotypic correlations from the NFBC1966 study (Appendix Table A5) as the generative covariance matrix Ω. We also assume a compound symmetry structure for the correlation matrix and vary the correlation ($\rho = 0.2$, 0.5, or 0.7) for the power analysis.

Type I Error and Power Analysis.: Similar to Simulation Study 1, the type I error rates are estimated in the presence and absence of environmental factor (Appendix Table A6). The power of these subset-based approaches are evaluated in two scenarios (Appendix Table A7) using the same metrics as used in Simulation Study 1. In Scenario 1, no interaction is involved, while in Scenario 2, interaction comes into play.

Data Analyses

We illustrate our methods by using data from two genetic association studies. The first is a meta-analysis of six independent case-control studies of T2D and the other considers a multiple-phenotype analysis of nine quantitative lipid-related traits measured on a

population-based cohort. In both examples, we compare the performance of pASTA with other competing methods as described in Table 1.

Application to Case-Control Studies of Type 2 Diabetes—This analysis includes data from a subset of subjects in six independent studies of T2D: FIN-D2D 2007 (D2D2007), The DIAbetes GENetic Study (DIAGEN), the Finland-United States Investigation of NIDDM Genetics Stage 2 (FUSION S2), The Nord-Trøndelag Health Study 2 (HUNT), the METabolic Syndrome In Men Study (METSIM) and the Tromsø Study (TROMSO) [25]. We investigated the associations of T2D and two SNPs in the *FTO* gene, rs6499640 and rs1121980, respectively after taking into account potential interaction effects with BMI. These two SNPs were chosen as candidates because they showed significant interactions with BMI in a previous study [26]. Descriptive statistics of the six studies are summarized in Appendix Table A8. In particular, the total sample size of the six studies was 9,624, with 4,422 cases and 5,202 controls. Sample sizes of the six studies vary from 1,058 to 2,219 and the case to control ratios vary from 0.37 to 1.75. The minor allele frequencies (MAF) of rs6499640 and rs1121980 in controls range from 0.37 to 0.43 and 0.39 to 0.48, respectively, across the studies.

The outcome of interest is the presence ($D = 1$) or absence ($D = 0$) of T2D. Each SNP is coded as $G = 0, 1, 2$, representing the number of minor alleles in rs6499640 or rs1121980 for each subject assuming an additive genetic susceptibility model. Let E denote BMI, A denote age in years, and S represent an indicator of sex (Male = 1). For the i-th individual in the k-th study, the standard logistic regression model incorporating G-E interaction is specified as

$$\log \frac{P(D_{ki} = 1 \mid E_{ki}, G_{ki}, Z_{ki})}{P(D_{ki} = 0 \mid E_{ki}, G_{ki}, Z_{ki})} = \beta_0^{(k)} + \beta_E^{(k)} \cdot \left(E_{ki} - \bar{E}_k\right) + \beta_G^{(k)} \cdot G_{ki}$$
$$+ \beta_{GE}^{(k)} \cdot G_{ki} \cdot \left(E_{ki} - \bar{E}_k\right) + \beta_A^{(k)} \cdot A_{ki} + \beta_S^{(k)} \cdot S_{ki} \qquad \text{(Model 3)}$$

and the standard logistic regression model of marginal SNP effect (adjusted for BMI, age and gender) is specified as

$$\log \frac{P(D_{ki} = 1 \mid E_{ki}, G_{ki}, Z_{ki})}{P(D_{ki} = 0 \mid E_{ki}, G_{ki}, Z_{ki})} = \alpha_0^{(k)} + \alpha_E^{(k)} \cdot \left(E_{ki} - \bar{E}_k\right) + \alpha_G^{(k)} \cdot G_{ki}$$
$$+ \alpha_A^{(k)} \cdot A_{ki} + \alpha_S^{(k)} \cdot S_{ki} . \qquad \text{(Model 4)}$$

Model 3 and Model 4 were fitted separately to the six studies. Then the various meta-analysis methods were applied to the summary statistics. Note that the METSIM study only contains male subjects, so the sex indicator was removed from the two models for this study.

Application to the North Finland Birth Cohort 1966 Study—We obtained the genotype and phenotype data of the NFBC1966 study from dbGaP with accession number phs000276.v2.p1. The NFBC1966 data contain 5,402 individuals and 364,590 SNPs, with multiple metabolic traits measured on subsets of individuals. Among these traits, we considered nine quantitative traits in the present study as phenotypes: C-reactive protein

(CRP), glucose, insulin, total cholesterol (TC), high density lipoprotein (HDL), low density lipoprotein (LDL), triglycerides (TG), systolic blood pressure (SBP), and diastolic blood pressure (DBP). Body mass index (BMI) was the environmental factor of interest. Following previous studies [27,28] we excluded individuals with missing phenotype data or having discrepancies between reported sex and sex determined from the X chromosome. We excluded SNPs with a minor allele frequency less than 1%, having missing values in more than 1% of the individuals, or with a Hardy-Weinberg equilibrium p-value below 10^{-4}. This left us with 5,123 individuals and 319,147 SNPs. For each phenotype in turn, we quantile transformed the phenotypic values to a standard normal distribution, regressed out sex, oral contraceptives and pregnancy status effects [27], and quantile transformed the residuals to a standard normal distribution again. We replaced the missing genotypes for a given SNP with its mean genotype value. To account for potential relatedness in the Finnish population one can use a kinship matrix in a linear mixed model to account for population admixture. As an alternative, we find that using linear regression models with principal components (PCs) as covariates also effectively controls for population stratification in the data. For example, with ten PCs, the genomic control factors for the ten phenotypes ranged from 0.96 to 1.01, with a median value of 0.99. Therefore, we extracted the top ten PCs from the genotype matrix as covariates to control for potential population stratifications.

The linear regression model with G-E interaction is specified as

$$Y_{ki} = \beta_0^{(k)} + \beta_G^{(k)} \cdot G_{ki} + \beta_E^{(k)} \cdot E_{ki} + \beta_{GE}^{(k)} \cdot G_{ki} \cdot E_{ki} + C_{ki}\beta_C^{(k)} \qquad \text{(Model 5)}$$

and the linear regression model of marginal SNP effect (adjusted for BMI and PCs) is specified as

$$Y_{ki} = \alpha_0^{(k)} + \alpha_G^{(k)} \cdot G_{ki} + \alpha_E^{(k)} \cdot E_{ki} + C_{ki}\alpha_C^{(k)} \qquad \text{(Model 6)}$$

where k indexes phenotypes, i indexes subjects, E denotes BMI, G denotes the number of minor alleles, and C denotes the vector of top ten PCs. We performed a genome-wide analysis of multiple phenotypes by fitting Model 5 and Model 6 separately to each of the nine phenotypes for all 319,147 SNPs in the data. Then for each SNP we calculated the meta-analytic p-values using ASSET and pASTA, respectively. Correlations among the resultant Z-statistics of the nine phenotypes were estimated using observed phenotypic correlations.

After the genome-wide screening using the nine phenotypes separately, we performed another genome-wide association analysis by analyzing all ten traits jointly with the multivariate linear mixed model implemented in GEMMA [29,30] and compared the results of the two GWAS. We then selected the top 100 SNPs that showed the strongest associations for subsequent interaction analysis. We performed SNP-BMI interaction analysis on these 100 SNPs.

Results

Simulation Study 1: Combining Signals Across Multiple Independent Case-Control Studies

Type I Error—In both scenarios, empirical type I error rates at thresholds $\alpha = 10^{-2}$, 10^{-3}, and 10^{-5} are well-preserved for all methods except pASTA(MA+EB) (Table 2), which tends to be conservative (e.g. type I error rate being 3.8×10^{-6} at the level of 10^{-5}). This has been noted in previous studies of EB-type adaptive methods [15].

Power Comparisons of ASSET and pASTA Based on 1-df Tests for Marginal Genetic Association—We extensively investigate how pASTA(MA) performs compared to ASSET when no environmental factors are considered (i.e. data generated under Model 1). The power of pASTA to detect overall genetic associations is almost the same as that of ASSET as the odds ratio of the marginal genetic effects varies, regardless of the directions of the genetic effects (Appendix Figures A1A, A1E). However, there is an observable difference in their accuracy of determining the exact subset of non-null studies for both scenarios (Appendix Figures A1B, A1F). When the marginal genetic effects of non-null studies exist in the same direction (in this case positive direction), the accuracy of pASTA to determine the correct subset increases as the signal grows stronger, and reaches an empirical power of 0.72 as the odds ratio goes up to 1.2. However, two-sided ASSET is rarely able to identify the exact true subset, and the corresponding power stays around zero as the odds ratio varies. The main reason for such low accuracy is that two-sided ASSET searches for associated subset in both positive and negative directions, identifies a subset that gives the largest $|Z(S)|$ for each direction, and takes the union of the two subsets as the final identified subset regardless of the significance of association in either direction. As a result, two-sided ASSET generally identifies more false positive studies when the significant associations are all in one direction. For example, when associations only exist in the positive direction in a subset of studies, the log odds ratios in the truly null studies are likely to be estimated as negative, and thus some null studies will be selected by two-sided ASSET. When the associations exist in opposite directions, two-sided ASSET has improved power to identify the correct subset of studies, but pASTA still consistently outperforms ASSET (Appendix Figures A1B, A1F). For both scenarios, the sensitivity of ASSET consistently stays higher than that of pASTA, while the specificity of ASSET is relatively low (Appendix Figures A1C, A1D, A1G, A1H). These results are intuitively reasonable, since sensitivity is the proportion of non-null studies that are correctly identified and does not require that the exact subsets to be identified.

When all of the ten studies are non-null (Appendix Figure A2A), pASTA is as powerful as ASSET in detecting signals of marginal associations, and only slightly less powerful than Fisher's method when the odds ratio of marginal genetic effects is small (<1.1). In addition, the fixed-effect meta-analysis method outperforms the other three 1-df methods (as expected) when the marginal odds ratio is less than 1.1. The power becomes very similar after this effect size. When only four out of the ten studies have true signals (Appendix Figure A2B), the subset-based methods provide power gain over the fixed-effect method.

Power Comparison of All Methods with Data Generated under Model 2

Detection of Association.: We consider five scenarios (Appendix Table A4) to evaluate the power of detecting overall genetic associations. When there exists only marginal genetic effect but no G-E interaction (Scenario 1), 1-df methods are slightly more powerful in discovering the overall associations than 2-df methods when the number of non-null studies $k_0 \leq 5$. For example, pASTA(JOINT) reaches a power of 0.85 when four out of the ten studies are non-null, which is 9.1%, 7.1%, and 9.9% lower than the power of Fisher's(MA), ASSET, and pASTA(MA), respectively (Figure 1A). This loss of power results from a penalty of the additional degree of freedom incorporated in the testing procedure in the absence of G-E interaction. In Scenario 2, when G-E interaction exists and marginal genetic effect is moderate, all 2-df methods provide substantial power gain over 1-df methods (Figure 1B). For example, when $k_0 = 7$, the power of pASTA(MA+EB) reaches 0.99, while that of ASSET stays around 0.6. In Scenario 3, where the marginal genetic effect is even smaller, the power of all methods is compromised, with pASTA(MA+EB) achieving the best power of 0.79 when $k_0 = 7$. The gold standard test has the second best performance with the power being 0.61 when $k_0 = 7$. All 1-df methods barely detect any signals (Figure 1C). When the marginal association is nearly null but sub-group effects of genetic factor present in opposite directions in the exposed and unexposed group (Scenario 4), 2-df methods are more powerful in detecting signals than their 1-df counterparts. Specifically, pASTA(MA +EB) performs the best when $k_0 > 3$, already achieving a power of 0.82 when $k_0 = 3$. Other 2-df methods perform slightly worse, with their power ranging from 0.44 to 0.52 when $k_0 = 3$, but they still yield power greater than 0.8 when $k_0 > 5$. All 1-df methods fail to detect the associations (Figure 1D). In a much less common situation where "cross-over" interaction exists (Scenario 5), the empirical power curves show a similar trend (Figure 1E) to those in Scenario 4. It is instructive to note that for Scenario 2–5 where the interaction effects are present, we tested the null hypothesis of JOINT for the gold standard test. Therefore, pASTA(MA+EB) outperforms the gold standard when k_0 is large since EB estimator exploits the putative G-E independence and can be more efficient in this case.

Subset Identification.: When the environmental factor is absent (Scenario 1), pASTA(MA) achieves the highest accuracy in identifying the exact subset of non-null studies, with the corresponding power being constantly around 0.4 (Figure 1F). The sensitivity of the subset-based methods, whether they are 1-df or 2-df, are close to one another as the number of non-null studies varies, and all of the methods are able to identify more than 80% of the non-null studies (Appendix Figure A3A). The specificity of ASSET is much smaller than that of the other subset-based methods based on pASTA (Appendix Figure A3F). When G-E interaction comes into play (Scenario 2–5), pASTA(MA+EB) performs the best in identifying the exact subset and has both the highest sensitivity and specificity across the four scenarios. In particular, in Scenario 4 where the genetic variant has opposite effects on exposed and unexposed groups, the highest empirical probability of pASTA(MA+EB) to select the correct non-null studies is 0.46 when three out of the ten studies are non-null. pASTA(JOINT) and pASTA(MA+CC) have the second-best performance in terms of determining the exact subset of non-null studies as well as sensitivity and specificity across Scenarios 2–5. It is noteworthy that in all five scenarios, the specificity of ASSET is the

lowest among the subset-based methods, which indicates that ASSET tends to give more false positives of non-null studies than pASTA.

Simulation Study 2: A Multiple-Phenotype Analysis in a Cohort Study

Type I Error—The results of the type I error rates are presented in Table 2. In general, the type I error rates of all methods tend to be conservative at all thresholds under evaluation. The only exception is that pASTA(JOINT) yields a type I error of 1.2×10^{-5} at the level of 10^{-5} when no environmental factor is considered. The conservativeness of the type I errors may be due to the inaccurate estimation of Σ, the correlation matrix of the Z-statistics.

Power Comparison

Detection of Association.: Figure 2A reveals that in the absence of interaction, pASTA(MA) outperforms the other two methods when over five out of the nine traits have genetic associations. The power of ASSET drastically falls when the number of non-null traits exceeds five due to computational issues of the ASSET package. According to Figure 2E, when interaction exists, pASTA(JOINT) provides great power gain in detecting the overall associations and reaches a power of 0.92 when five out of the nine traits are non-null. The power of the 1-df methods stays below 0.9. We also investigate the relative power of the subset-based methods with respect to different correlation values among phenotypes. low (ρ=0.2), moderate (ρ=0.5), or high (ρ=0.7). We removed ASSET from comparison due to the non-convergence of its algorithm when phenotypic correlations are large. Appendix Figures A4A and A4D show that pASTA has high power power when the correlation between phenotypes is low (ρ=0.2). For example, when G-E interaction exists, pASTA(JOINT) achieves a power of 0.8 when five out of the nine traits are non-null (Appendix Figure A4D). When the correlation becomes moderate (ρ=0.5) and large (ρ=0.7), the power of pASTA reduces, with pASTA(JOINT) achieving power around 0.60 and 0.45, respectively, in the presence of G-E interaction when seven out of the nine traits are non-null (Appendix Figures A4E, A4F).

Subset Identification.: Similar to what we observe in Simulation Study 1, the accuracy of ASSET in identifying the exact subset of non-null traits are constantly close to 0 in both scenarios (Figures 2B, 2F). In Figures 2C and 2G, the sensitivity of the pASTA approaches is close to one another and stays above 0.8 as the number of non-null traits varies, which indicates that these methods are able to identify over 80% of the non-null traits in one study. On the other hand, the sensitivity of ASSET decreases as the number of non-null traits increases. The curves of specificity (Figures 2D, 2H) show similar trends to those of sensitivity.

Application 1: Type 2 Diabetes Data

The exponential of the parameter estimates ($\alpha_G^{(k)}$, $\beta_G^{(k)}$ and $\beta_{GE}^{(k)}$), their corresponding 95% confidence interval (CI) and p-values obtained from Model 3 and Model 4 for each of the six studies are summarized in Appendix Table A9. The EB estimates of the interaction terms and their associated Wald chi-square statistics were obtained using the R package CGEN [31]. For SNP rs6499640, the odds ratios of marginal genetic effect range from 0.89 to 1.11,

and none of them are significant at the level of 0.05. When the interaction between SNP and BMI is included in the model, the odds ratios of genetic main effect do not change remarkably from those of marginal genetic effects. However, despite the small effect sizes of interactions in all six studies, the interaction effects are statistically significant in D2D2007 for CC method (p=0.017) and in FUSION S2 for CC (p=0.0006) and EB (p=0.003) methods. The SNP rs1121980 has significant marginal associations with T2D in FUSION S2 (OR_G = 1.19, p=0.011) and HUNT (OR_G = 1.21, p=0.020), while none of the studies show significant SNP-BMI interactions for this SNP.

We then meta-analyzed the results from the six studies by applying all methods in Table 1, and we present the results of the eight methods in Table 3. For SNP rs6499640, all 1-df methods give insignificant p-values for marginal genetic associations with T2D at the level of 0.05. As for 2-df methods, pASTA(JOINT) gives the smallest p-value (p=0.008), and pASTA(MA+CC) give a slightly higher p-value (p=0.009). The result of pASTA(MA+EB) is not statistically significant (p=0.090), which may be due to the presence of G-E correlation in this example. All 2-df pASTA methods identify D2D2007 and FUSION S2 as the non-null studies, which is consistent with what we observe from the results of 2-df tests for each study separately. We also investigated the odds ratios of T2D across different levels of BMI by fitting a logistic regression adjusting for age, gender, and the studies to which the subjects belong. Figure A4 in Appendix shows that the marginal odds ratio for the SNP rs6499640 is close to 1, but when G-E interaction is taken into account, this SNP reveals a significant protective effect at the level of 0.05 when BMI equals 20 and 25. When BMI equals 30, the odds ratio of T2D becomes positive and is significantly different from the one when BMI equals 20.

When there is no evidence for G-E interactions in individual studies, as with SNP rs1121980, all 2-df methods fail to detect the overall genetic associations, but the p-values obtained from 2-df pASTA are smaller than those from 2-df Fisher's p-value combined methods. Furthermore, the p-value given by pASTA(MA) (p=0.024) is smaller than those of Fisher's(MA) (p=0.080) and ASSET (p=0.031). pASTA also provides more plausible subset identification than ASSET when there are only significant marginal genetic effects. In addition to FUSION S2 and HUNT as identified by all pASTA methods, ASSET also includes D2D2007, METSIM, and TROMSO as associated studies, whose p-values with respect to marginal associations in study-specific analysis (Appendix Table A9) are far from significant. This data analysis demonstrates that without sacrificing much efficiency in the absence of G-E interaction, pASTA provides potential power gain when marginal genetic effects are modest and interactions are involved. Moreover, pASTA yields a more plausible subset of non-null studies.

Application 2: North Finland Birth Cohort 1966 Data

We performed a genome-wide multiple-phenotype analysis by applying pASTA and ASSET to the lipid-related traits in the NFBC1966 study. Appendix Table A10 presents the SNPs whose meta-analytic p-values are smaller than the genome-wide significance level ($p = 5 \times 10^{-8}$). Seventeen SNPs are identified by pASTA(MA) at the genome-wide significance level. ASSET identifies seven SNPs as the genome-wide significant SNPs, and the only

overlapping SNP between the two sets is rs754524, which is located in the *APOB* region on chromosome 2. However, it should be noted that since ASSET fails to produce p-values for some of the SNPs due to computational issues, the list of genome-wide significant SNPs generated by ASSET is not comprehensive and thus is not comparable to the list generated by pASTA(MA). Taking G-E interaction into consideration, pASTA(JOINT) identifies 13 SNPs, which are a subset of the SNPs identified by pASTA(MA). The SNP rs3764261, located on chromosome 16 near the *CETP* gene, is identified by pASTA(MA) ($p = 3.6 \times 10^{-20}$) and pASTA(JOINT) ($p = 7.1 \times 10^{-20}$) as the top SNP. Figure 3 presents the QQ plots corresponding to the three methods.

We then compared the 17 genome-wide significant SNPs identified by pASTA(MA) to the 23 SNPs identified by another GWAS where all ten traits were analyzed jointly using a multivariate marginal association test (i.e. GEMMA) (Appendix Table A10). Seven SNPs belong to both lists, and the top five SNPs identified respectively by GEMMA and pASTA(MA) are the same (though in slightly different order). We considered the top 100 SNPs identified by GEMMA for follow-up interaction analysis. The p-values produced by GEMMA for the 100 SNPs range from 5.5×10^{-30} to 7.0×10^{-5}. The results of the combined p-values for the complete set is presented in Appendix Table A11. Two SNPs, rs2083637 (on chromosome 8, near gene *LPL*) and rs754524 (on chromosome 2, in the *APOB* region), are used to illustrate the performance of the methods in the presence (rs2083637) or absence (rs754524) of G-E interaction. These two examples offer insight into the properties of the proposed methods and illustrate the advantage of incorporating G-E interaction into the analytic framework very clearly.

Appendix Table A12 summarizes the linear regression coefficients and their corresponding 95% CI and p-values obtained from Model 5 and Model 6. The SNP rs2083637, located near the *LPL* gene, is marginally associated with HDL and TG, with coefficients being −0.1 ($p = 3.4 \times 10^{-6}$) and 0.1 ($p = 5.6 \times 10^{-6}$), respectively. CRP ($p = 1.3 \times 10^{-3}$), TC ($p = 6.3 \times 10^{-3}$), and LDL ($p = 8.2 \times 10^{-3}$) are associated with rs2083637 through SNP-BMI interactions. On the other hand, rs754524 has a positive marginal effect on HDL ($p = 3.6 \times 10^{-2}$) and a negative marginal effect on TC ($p = 6.3 \times 10^{-10}$), LDL ($p = 2.1 \times 10^{-11}$), and TG ($p = 1.8 \times 10^{-2}$), while no significant SNP-BMI interactions are observed for these nine phenotypes under the 0.05 threshold.

For rs2083637, ASSET and pASTA(MA) both identify TG and HDL as associated phenotypes (Table 4). However, both pASTA(JOINT) and pASTA(MA+CC) identify three additional phenotypes (CRP, TC, LDL), for which rs2083637 has significant interaction with BMI, when G-E interaction is taken into account. Appendix Figure A5 presents the effect and the corresponding 95% CIs for rs2083637 on CRP, TC, and LDL when BMI equals the mean, mean±1 standard deviation (sd), and mean±2sd. We observe variation of the SNP effect sizes across different levels of BMI. For example, rs2083637 has positive but not significant effect on CRP for subjects with average BMI, while for subjects whose BMI is 2 sd greater than the average, the association becomes negative and significant at the level of 0.05. Therefore, the incorporation of G-E interaction can enhance the knowledge that is not available in marginal association testing. The overall p-value of associations given by pASTA(JOINT) (2.5×10^{-6}) is also smaller than those given by 1-df methods (2.7×10^{-5} for

pASTA[MA] and 3.6×10^{-6} for ASSET). For rs754524, pASTA(MA) identifies two phenotypes (TC and LDL), both of which have significant marginal associations with this SNP in phenotype-specific analysis, as associated phenotypes. SNP rs754524 has been previously reported to be associated with both traits. [32,33]. ASSET identifies two additional phenotypes, glucose and HDL, and fails to identify TC. The evidence for the marginal association of this SNP with HDL is not strong [28], which indicates that such identified association may be false. pASTA (JOINT) identifies the same set of associated traits as pASTA(MA). This example shows that by incorporating interactions one can obtain smaller p-values for detecting overall associations and discover more relevant traits that are missed by ASSET, ones that show significant G-E interaction effects in phenotype-specific analysis.

Discussion

We propose a powerful subset-based framework for the analysis of association studies after accounting for potential G-E interactions. Simulation studies demonstrate that this framework improves the power to detect overall associations as well as the accuracy of determining associated studies/traits by incorporating G-E interactions, while maintaining comparable power to ASSET in the absence of such interactions. Our data examples exemplify that 2-df pASTA yields smaller meta-analytic p-values for SNPs that show significant G-E interactions in study/phenotype-specific analysis compared to 1-df methods where only marginal genetic associations are considered (Table 3 and Table 4). In addition, our analysis on the NFBC1966 data illustrates that pASTA is able to identify new phenotypes that are associated with given susceptibility loci after considering G-E interactions that will be missed by traditional marginal association analysis (Table 4). These properties make pASTA a powerful tool to identify candidate SNPs that are potentially associated with the trait(s) of interest in the screening step of a GWAS, which then can be followed by deeper characterization/interpretation of the associations through other methods, such as biological validation. For example, RNA sequencing experiments can be performed to examine whether the candidate SNPs affect the expression level of the target gene [34]. Crispr-Cas9 knockout screening can be carried out in human cell lines to determine the functional significance of these candidate SNPs [35].

Our simulation studies show that in the presence of interaction, pASTA(MA+EB) has the best performance in terms of all four evaluation metrics in a case-control setting where G-E independence holds. The rest of the 2-df methods (Fisher's methods, pASTA[JOINT] and pASTA[MA+CC]) give similar power for detecting overall associations across the scenarios (Figures 1B–1E), but pASTA has the advantage of being able to identify the subset of studies/traits that are most likely to contribute to the associations. For a cohort study with multiple phenotypes where the G-E independence does not apply, pASTA(JOINT) gives the best performance in terms of the four metrics (Figures 2E–2H).

There are several limitations of our proposed approach that lend the problem to future research. The present proposal for pASTA focuses on testing the existence of genetic associations using p-values as inputs and is purely bi-directional in nature. It aggregates the evidence of both genetic effects and G-E interactions from individual studies but does not

specify the directions of such effects or interactions. It is a pure testing approach and does not provide pooled regression coefficient estimates from meta-analysis. Thus it loses insight into the nature of the environmental heterogeneity across cohorts. Further work is needed to define the subsets of studies that have similar direction of genetic effects in subsets defined by the environmental factor (for example the sign of estimates β_G in the group $E = 0$ and of $\beta_{GE} + \beta_G$ in the group $E = 1$ for a binary exposure E). The method is currently based on the assumption of fixed effect(s) meta-analysis. Since the input is p-values, the method can be extended to random effects meta-analysis as well as hybrid methods that combine fixed and random effects meta-analysis [4]. In our simulation studies and data analysis, pASTA uses phenotypic correlations to approximate the correlations among Z_k's, which may result in compromised power and conservative type I error. In addition, such approximation requires individual level data on multiple traits, which are not always available in GWAS. In the case where only summary data are accessible, an alternative is to follow the idea in Liu and Lin [36] and estimate the phenotypic correlations using the summary across a large number of independent null SNPs, since the phenotypic correlations are independent of genotypes under the null for a 1-df Wald test. However, using phenotypic correlations in place of the Z-statistics correlations when our proposed Z-statistics are obtained by transforming p-values of 2-df tests is only an ad hoc choice and by no means optimal. Despite the fact that our choice of using phenotypic correlations is a convenient approximation that is not theoretically proven, our simulation studies show that the properties of the resultant testing procedures are desirable. Optimizing the estimation of such correlation beyond 1-df Wald tests is an issue that needs to be further addressed.

There are some existing methods that incorporate phenotypic correlations into the construction of test statistics (as opposed to the evaluation of p-values in our method). For example, Dutta et al. [37] have studied the effect of modeling kernel matrix in the score test statistic using estimated phenotype covariance for testing pleiotropic effects of rare variants based on multiple kernel regression (Multi-SKAT). While we only use phenotypic correlation for characterizing the distribution of our combined statistics instead of using it directly during the process of constructing the statistics, incorporating phenotypic correlation into the test statistics may further improve the power of our method.

We only focused on candidate loci with known marginal association or gene-environment interaction in our data examples. Simulation results indicate that pASTA is scalable to a genome-wide level and maintains Type 1 error at smaller 10^{-5} threshold of α. In the T2D data analysis, the average run time of pASTA for the marginal effect, i.e. pASTA(MA), for one SNP was 181.7 milliseconds on a 2.9 GHz Intel processor, and the incorporation of G-E interaction, i.e. pASTA(JOINT), yielded a similar computation time (159.2 milliseconds). For the data from the NFBC 1966 study where correlations across the test statistics need to be taken into account, the average run times for both pASTA(MA) and pASTA(JOINT) for one SNP were 1.7 seconds. For a typical GWAS to analyze one million SNPs, the projected computation time of pASTA on 20 CPU cores is approximately one day. ASSET was slightly faster than pASTA in the NFBC 1966 data analysis (1.5 seconds) but twice slower in the T2D data analysis (3.3 seconds). Computing time was calculated using the R package microbenchmark [38]. We plan to optimize the pASTA implementation code and develop an efficient R package akin to ASSET.

Acknowledgement

Appendix

Detailed Derivation of Meta-analytic p-value

For an observed test statistic $Z_{meta} = T_{obs}$, the DLM approximated p-value can be expressed as

$$
\begin{aligned}
P(Z_{meta} > T_{obs}) &= P(\cup_{S \in \mathcal{S}} \{Z(S) > T_{obs}\}) \\
&\leq P(\cup_{\gamma \in \Gamma} \{Z(S_\gamma) > T_{obs} \text{ and } Z(S_\gamma) \text{ being a local maxima}\}) \\
&\leq \sum_{\gamma \in \Gamma} P(Z(S_\gamma) > T_{obs} \text{ and } Z(S_\gamma) \text{ being a local maxima}).
\end{aligned}
$$

For a fixed subset $S_\gamma \subseteq \{1, \cdots, K\}$,

$$
\begin{aligned}
&P(Z(S_\gamma) > T_{obs} \text{ and } Z(S_\gamma) \text{ being a local maxima}) \\
&= \int_{T_{obs}}^{\infty} P(Z(S_\gamma) \text{ being a local maxima} \mid Z(S_\gamma) = z)h(z)dz \\
&= \int_{T_{obs}}^{\infty} P(Z(S_\gamma \pm k) < z, \forall k \mid Z(S_\gamma) = z)h(z)dz \quad \text{(Equation A1)} \\
&\leq \int_{T_{obs}}^{\infty} \prod_{k=1}^{K} P(Z(S_\gamma \pm k) < z \mid Z(S_\gamma) = z)h(z)dz,
\end{aligned}
$$

where $h(\cdot)$ is the probability distribution function of $Z(S_\gamma)$. The last inequality is justified by the "separability" assumption that given the current subset, its neighbors have independent Z-statistics.

    i.   When $k \notin S_\gamma$, given that $Z(S_\gamma) = z$,

$$
\begin{aligned}
Z(S_\gamma + k) &= \sum_{j \in S_\gamma} \sqrt{\frac{n_j}{\sum_{l \in S_\gamma} n_l + n_k}} \cdot Z_j + \sqrt{\frac{n_k}{\sum_{l \in S_\gamma} n_l + n_k}} \cdot Z_k \\
&= \left(\sum_{j \in S_\gamma} \sqrt{\frac{n_j}{\sum_{l \in S_\gamma} n_l}} \cdot Z_j \right) \cdot \sqrt{\frac{\sum_{l \in S_\gamma} n_l}{\sum_{l \in S_\gamma} n_l + n_k}} + \sqrt{\frac{n_k}{\sum_{l \in S_\gamma} n_l + n_k}} \cdot Z_k \\
&= z \cdot \sqrt{\frac{\sum_{l \in S_\gamma} n_l}{\sum_{l \in S_\gamma} n_l + n_k}} + \sqrt{\frac{n_k}{\sum_{l \in S_\gamma} n_l + n_k}} \cdot Z_k.
\end{aligned}
$$

    Hence,

$$Z(S_\gamma + k) < z \Leftrightarrow Z_k < \frac{\left(1 - \sqrt{\frac{\sum_{l \in S_\gamma} n_l}{\sum_{l \in S_\gamma} n_l + n_k}}\right) \cdot z}{\sqrt{\frac{n_k}{\sum_{l \in S_\gamma} n_l + n_k}}} \equiv u_k(z) \qquad \text{(Equation A2)}$$

and

$$P(Z(S_\gamma + k) < z \,|\, Z(S_\gamma) = z) = P(Z_k < u_k(z) \,|\, Z(S_\gamma) = z)$$

ii. When $k \in S_\gamma$, given that $Z(S_\gamma) = z$,

$$Z(S_\gamma - k) = \sum_{j \in S_\gamma - \{k\}} \sqrt{\frac{n_j}{\sum_{l \in S_\gamma - \{k\}} n_l}} \cdot Z_j$$

$$= \left(\sum_{j \in S_\gamma} \sqrt{\frac{n_j}{\sum_{l \in S_\gamma} n_l}} \cdot Z_j\right) \cdot \sqrt{\frac{\sum_{l \in S_\gamma} n_l}{\sum_{l \in S_\gamma - \{k\}} n_l}} - \sqrt{\frac{n_k}{\sum_{l \in S_\gamma - \{k\}} n_l}} \cdot Z_k$$

$$= z \cdot \sqrt{\frac{\sum_{l \in S_\gamma} n_l}{\sum_{l \in S_\gamma - \{k\}} n_l}} - \sqrt{\frac{n_k}{\sum_{l \in S_\gamma - \{k\}} n_l}} \cdot Z_k.$$

Hence,

$$Z(S_\gamma - k) < z \Leftrightarrow Z_k < \frac{\left(\sqrt{\frac{\sum_{l \in S_\gamma} n_l}{\sum_{l \in S_\gamma} n_l + n_k}} - 1\right) \cdot z}{\sqrt{\frac{n_k}{\sum_{l \in S_\gamma} n_l + n_k}}} \equiv l_k(z) \qquad \text{(Equation A3)}$$

$$P(Z(S_\gamma - k) < z \,|\, Z(S_\gamma) = z) = P(Z_k > l_k(z) \,|\, Z(S_\gamma) = z).$$

This conditional probability of Z_k given $Z(S_\gamma)$ can be evaluated based on a joint bivariate normal distribution, which is specified in the main text. For numerical evaluation of the integrals, we directly apply the R function integrate().

Metrics Used in Power Analysis

We used two main metrics for evaluation:

- **Power of detecting genetic associations that truly exist**, defined as the probability of rejecting the relevant null hypotheses H_{0k} within study k with a pre-specified level a. This quantity is estimated as the empirical proportion of replications where the meta-analytic p-value combined across all studies is smaller than a;

- **Power of identifying the exact true subset of non-null studies**, defined as the probability of correctly identifying such subset. The definition of non-null studies is specified above. The corresponding empirical proportion of replications declaring the correct subset to be non-null is used to estimate this power.

We also report the sensitivity and specificity of identifying the subset of non-null studies, defined as the proportion of non-null studies that are correctly identified by subset-based approaches, and the proportion of truly null studies that are declared null, respectively. They are both estimated by the corresponding empirical proportions in the simulation. Note that sensitivity and specificity respectively allow false positives and false negatives in the identified subset, while power of identifying exactly the true subset is a much more stringent criterion.
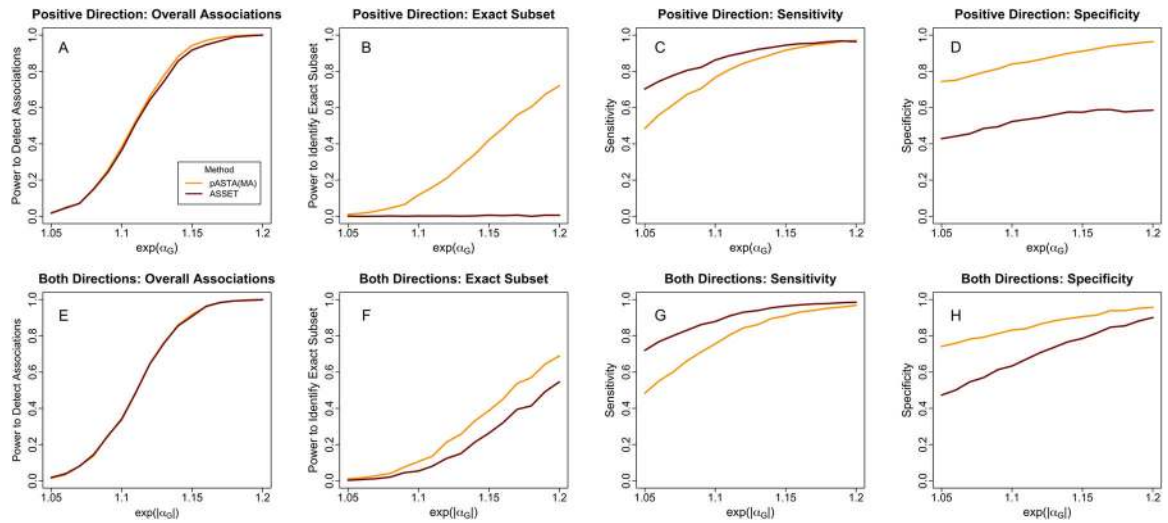
Figure A1. Power Comparisons of Two-Sided ASSET and pASTA for Detecting Marginal Genetic Associations Using 1-df Tests.

In each setting, ten studies are considered, with 6,000 cases and 6,000 controls in each study. Four of the studies contain true marginal genetic associations. The upper panels assume that all associations are in the positive direction; the lower panels assume that 50% of the associations are positive and the remaining 50% are negative.

Overall Associations
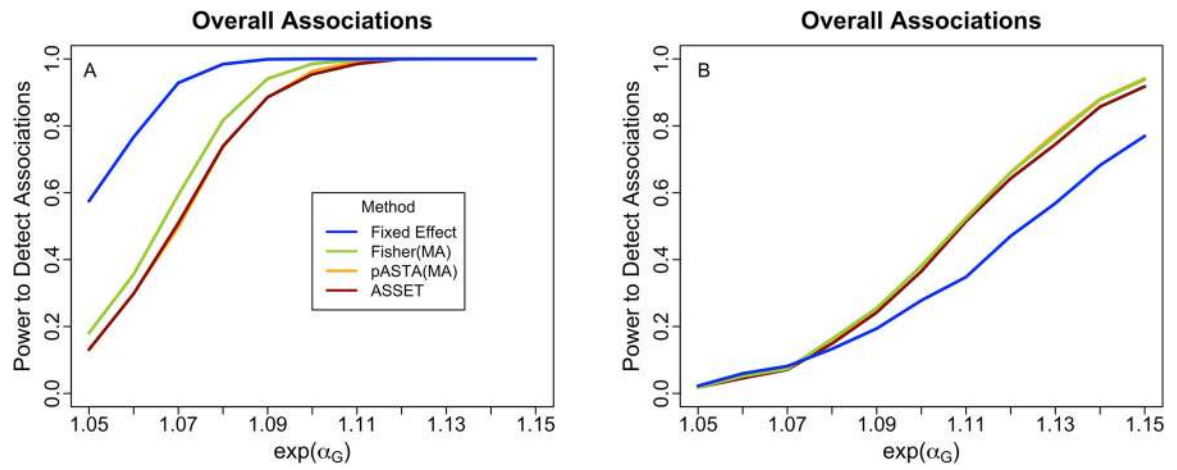
Figure A2. Empirical Power Curves for Detecting Marginal Genetic Associations When (A) All 10 Studies Are Non-Null and (B) 4 out of the 10 Studies Are Non-Null.
Ten studies are considered, with 6,000 cases and 6,000 controls in each study. Simulations are repeated 2,000 times for each odds ratio.
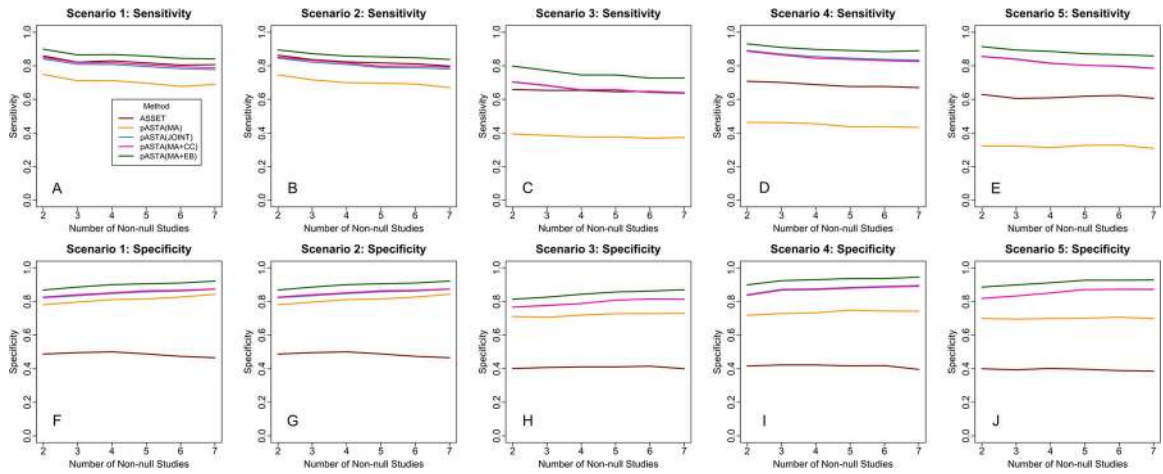
Figure A3. Sensitivity and Specificity of Nine Meta-Analysis Methods in Case-Control Studies

Sensitivity (top row) and specificity curves (bottom row) are plotted only for subset-based meta-analysis methods ASSET and pASTA. In each simulation, a total of ten studies are analyzed, and each study contains 6,000 cases and 6,000 controls. Scenario 1 considers no environmental main effect or interaction, representing a situation where only marginal genetic effects exist (data generated from Model 1); Scenario 2 assumes strong interaction and moderate genetic main effect; Scenario 3 assumes strong interaction and no genetic main effect; Scenario 4 assumes positive interactions and negative genetic main effects, resulting in null marginal genetic effects; Scenario 5 considers negative interaction and positive genetic main effect.
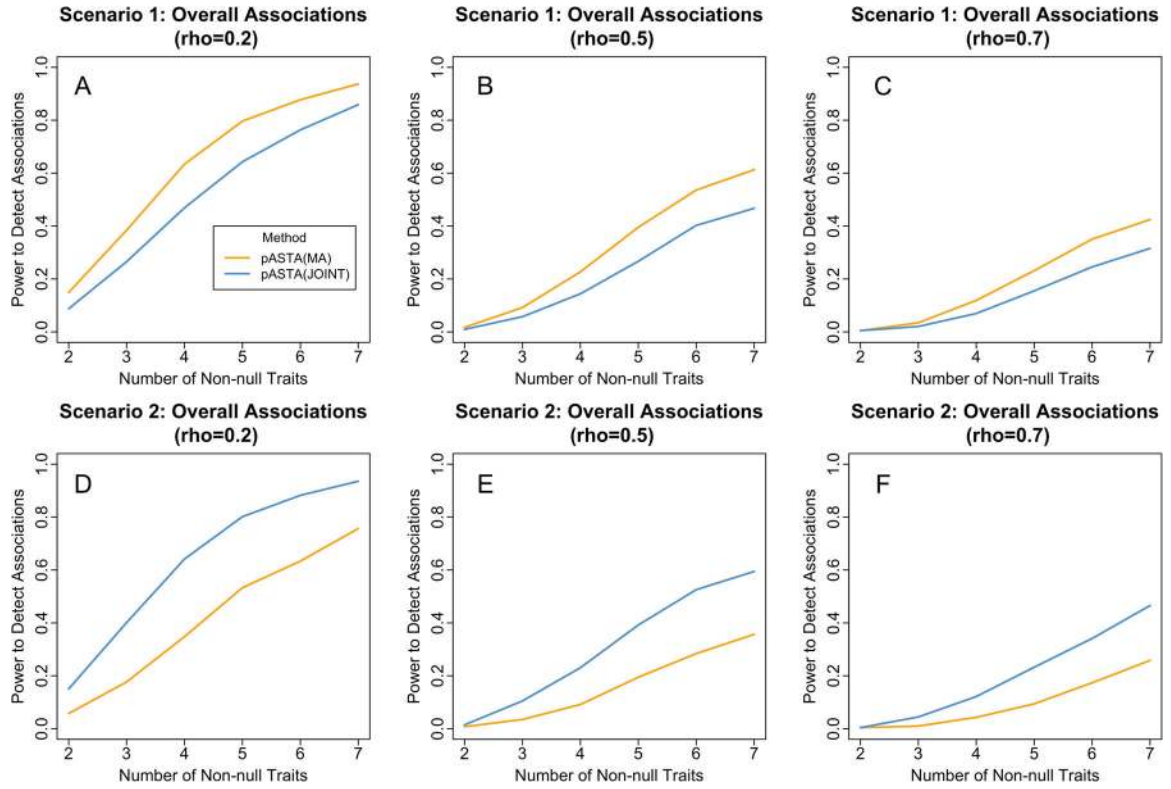
Figure A4. Power of Subset-Based Methods for Detecting Overall Associations with Varying Phenotypic Correlations.

A cohort study with nine correlated phenotypes measured and 6,000 subjects are simulated. The performance of the pASTA is evaluated in the absence (top row) and presence (bottom row) of gene-environment interaction. A compound symmetry correlation structure is assumed. The left, middle, and right column represent low, moderate, high correlation among phenotypes, respectively. The level of significance is 0.001, and simulations are repeated 2,000 times.
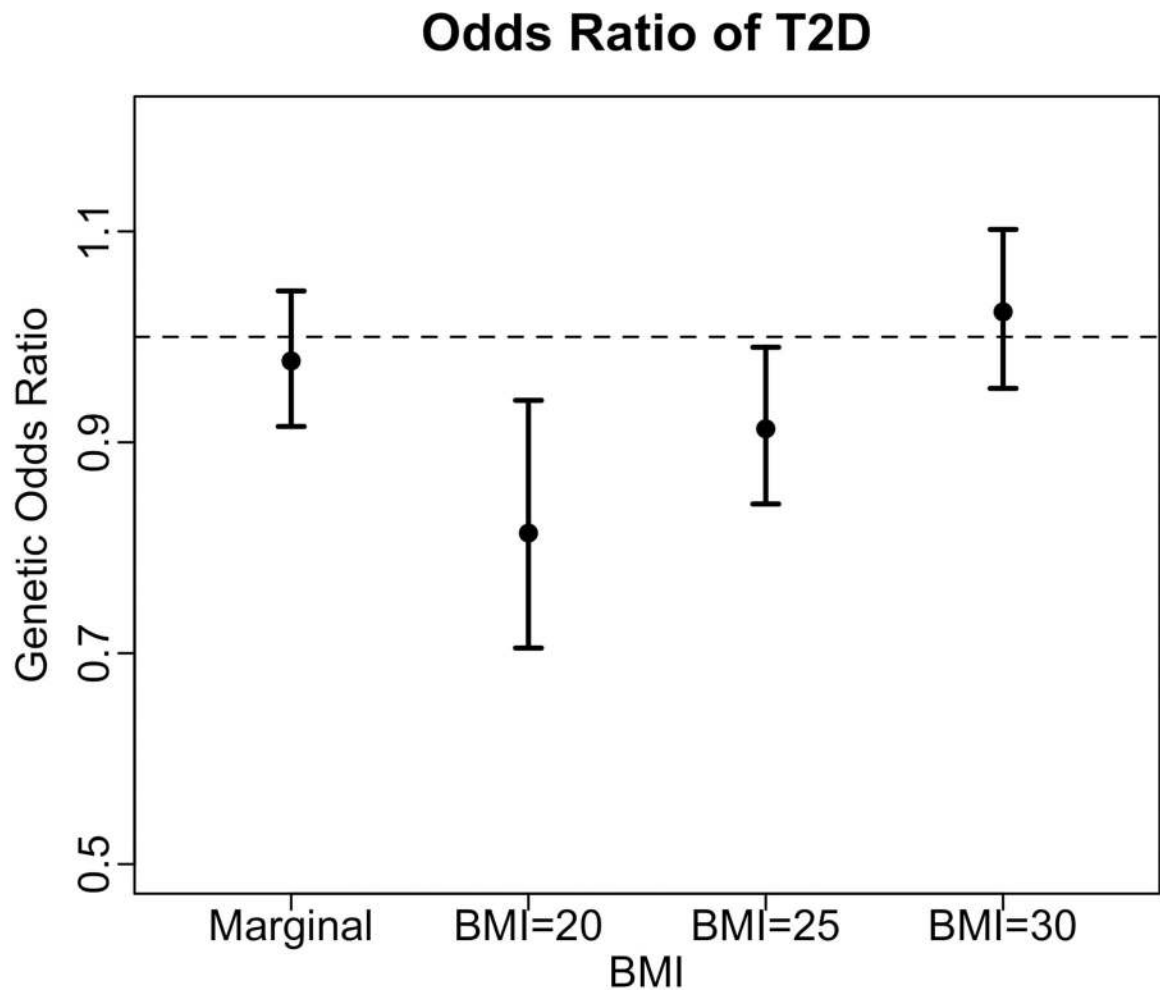
Odds Ratio of T2D

Figure A5. Odds Ratios of T2D and Associated 95% Confidence Intervals for SNP rs6499640 for Various Levels of BMI.
The SNP odds ratios of T2D are estimated by fitting a logistic regression model adjusted for age, gender, and the studies (e.g. METSIM) to which the subjects belong.
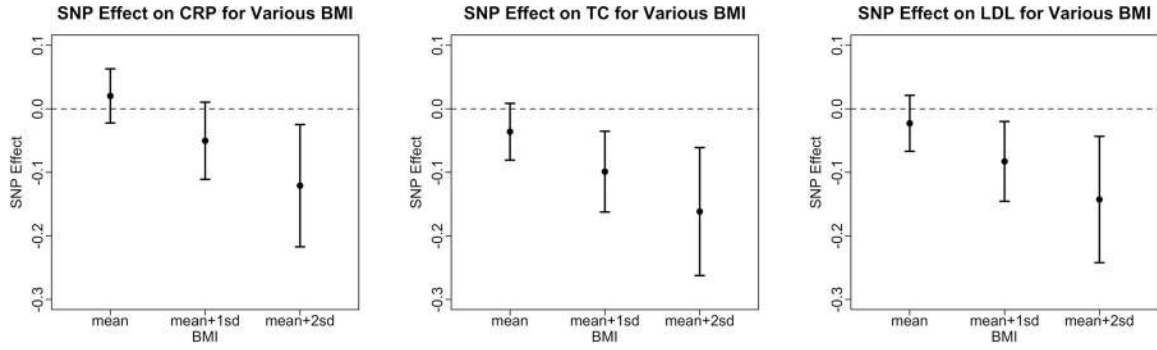
Figure A6. Effect of the SNP rs2083637 on CRP, TC, and LDL for Various Levels of BMI and Associated 95% Confidence Intervals.
SNP effect is estimated using data from the North Finland Birth Cohort 1966 Study. Dashed line indicates zero effect.

Abbreviations: CRP, C-reactive Protein; LDL, Low Density Lipoprotein; TC, Total Cholesterol.

Table A1.

Types of Association Tests Applied to a Single

Types of Association Test	Null Hypotheses (H_{0k})	Test Statistics	Distributions under Null	p-values
MA	$\alpha_G^{(k)} = 0$ [†]	$\chi^2_{(k)} = \dfrac{\left(\hat{\alpha}_G^{(k)}\right)^2}{\widehat{Var}\left(\hat{\alpha}_G^{(k)}\right)}$	Chi-square distribution with 1 df $\left(\chi_1^2\right)$	$p_k = P\left(\chi_1^2 > \chi^2_{(k)}\right)$
JOINT	$\beta_G^{(k)} = 0,$ $\beta_{GE}^{(k)} = 0$	$\chi^2_{(k)} = -2(\ell_0 - \ell_1)$ [‡]	Chi-square distribution with 2 df $\left(\chi_2^2\right)$	$p_k = P\left(\chi_2^2 > \chi^2_{(k)}\right)$
MA+GE [§]	$\alpha_G^{(k)} = 0,$ $\beta_{GE}^{(k)} = 0$	$\chi^2_{(k)} = \dfrac{\left(\hat{\alpha}_G^{(k)}\right)^2}{\widehat{Var}\left(\hat{\alpha}_G^{(k)}\right)} + \dfrac{\left(\hat{\beta}_{GE}^{(k)}\right)^2}{\widehat{Var}\left(\hat{\beta}_{GE}^{(k)}\right)}$	Chi-square distribution with 2 df $\left(\chi_2^2\right)$	$p_k = P\left(\chi_2^2 > \chi^2_{(k)}\right)$

[†] $\alpha_G^{(k)}$ is the coefficient of marginal genetic effect in Model 1, while $\beta_G^{(k)}$ and $\beta_{GE}^{(k)}$ are coefficients of genetic main effect and G-E interaction, respectively, in Model 2.

[‡] ℓ_0 is the log-likelihood under the null hypothesis and ℓ_1 is the unconstrained log-likelihood.

[§] The interaction coefficient $\beta_{GE}^{(k)}$ can be estimated by case-control $\left(\hat{\beta}_{CC}^{(k)}\right)$ or empirical Bayes $\left(\hat{\beta}_{EB}^{(k)}\right)$ estimator in a case-control study.

Table A2.

Comparison of ASSET and pASTA

	ASSET	pASTA
Study-specific Z-statistic	$Z_k^A = \hat{\alpha}_G^{(k)}/s_k$	$Z_k^P = -\Phi^{-1}(p_k)$

	ASSET	pASTA
Subset Z-statistic	$Z^A(S) = \sum_{k \in S} w_k Z_k^A$	$Z^P(S) = \sum_{k \in S} w_k Z_k^P$
Meta-analysis test statistic	$Z_{meta}^A = \max_{S \in \mathscr{S}} \left| Z^A(S) \right|$	$Z_{meta}^P = \max_{S \in \mathscr{S}} Z^P(S)$
Subset identified	Subset that gives the largest $|Z^A(S)|$	Subset that gives the largest $Z^P(S)$
DLM-based p-value	$p_{DLM} = \sum_{S \in \mathscr{S}} \int_T^\infty 2 \cdot \prod_k P\big(Z^A(S \pm k) < z \,|\, Z^A(S) = z\big) h(z) dz$	$p_{DLM} = \sum_{S \in \mathscr{S}} \int_T^\infty \prod_k P\big(Z^P(S \pm k) < z \,|\,$

Table A3.

Parameter Settings for the Evaluation of Type I Error Rates for Independent Studies

Parameters	θ_{GE}	p_G	p_E	β_{GE}	β_G	β_E	n_0	n_1	OR_G	OR_E
Scenario 1	0	0.2	0.3	0	0	0	6,000	6,000	1.0	1.0
Scenario 2	0	0.2	0.3	0	0	log 1.4	6,000	6,000	1.0	1.4

θ_{GE}, p_G and p_E denote the G-E associations, the frequency of $G = 1$ and the prevalence of environmental factor ($E = 1$) in controls, respectively. OR_G and and OR_E respectively represent the odds ratios of marginal genetic and environmental effects.

Table A4.

Parameter Settings for Power Analysis for Independent Studies

Parameters	p_G	p_E	β_{GE}	β_G	β_E	OR_G	OR_E
Scenario 1	2	0.3	0	log 1.15	0	1.15	0
Scenario 2	0.2	0.1	log 1.3	log 1.05	log 1.2	1.09	1.27
Scenario 3	0.2	0.1	log 1.3	0	0	1.03	1.06
Scenario 4	0.2	0.3	log 1.3	−log 1.15	log 1.2	0.96	1.26
Scenario 5	0.2	0.3	−log 1.3	log 1.1	log 1.2	1.01	1.14

p_G and p_E denote the frequency of $G = 1$ and the prevalence of environmental factor ($E = 1$) in controls, OR_G and OR_E respectively represent the odds ratios of marginal genetic and environmental effects.

Table A5.

Phenotypic Correlations of the Data from North Finland Birth Cohort 1966 Study

	CRP	Glucose	Insulin	TC	HDL	LDL	TG	SBP	DBP
CRP	1	0.07	−0.09	0.08	−0.22	0.11	0.24	0.08	0.07
Glucose	0.07	1	−0.04	0.03	−0.07	0.02	0.13	0.15	0.07
Insulin	−0.09	−0.04	1	−0.03	0.08	−0.02	−0.09	−0.04	−0.01
TC	0.08	0.03	−0.03	1	0.13	0.88	0.41	0.11	0.10
HDL	−0.22	−0.07	0.08	0.13	1	−0.14	−0.40	−0.03	−0.04
LDL	0.11	0.02	−0.02	0.88	−0.14	1	0.33	0.09	0.09
TG	0.24	0.13	−0.09	0.41	−0.40	0.33	1	0.15	0.12

	CRP	Glucose	Insulin	TC	HDL	LDL	TG	SBP	DBP
SBP	0.08	0.15	−0.04	0.11	−0.03	0.09	0.15	1	0.30
DBP	0.07	0.07	−0.01	0.10	−0.04	0.09	0.12	0.30	1

Abbreviations: CRP, C-reactive Protein; TC, Total Cholesterol; HDL, High Density Lipoprotein; LDL, Low Density Lipoprotein; TG, Triglycerides; Systolic Blood Pressure (SBP); Diastolic Blood Pressure (DBP).

Table A6.

Parameter Settings for the Evaluation of Type I Error Rates for Multiple Correlated Phenotypes

Parameters	p_G	p_E	β_{GE}	β_G	β_E	n
Scenario 1	0.2	0.3	0		0	6,000
Scenario 2	0.2	0.3	0	0	0.3	6,000

p_G and p_E denote the frequency of $G = 1$ and the prevalence of environmental factor ($E = 1$) respectively. β_G, β_E, and β_G are coefficients of genetic main effect, environmental main effect, and gene-environment interaction in Model 2 under an identity link, respectively. n denotes the sample size of the study.

Table A7.

Parameter Settings for Power Analysis for Multiple Correlated Phenotypes

Parameters	p_G	p_E	β_{GE}	β_G	β_E	n
Scenario 1	0.2	0.3	0	0.1	0.15	6,000
Scenario 2	0.2	0.3	0.15	0.04	0.15	6,000

p_G and p_E denote the frequency of $G = 1$ and the prevalence of environmental factor ($E = 1$) respectively. β_G, β_E, and β_G are coefficients of genetic main effect, environmental main effect, and gene-environment interaction in Model 2 under an identity link, respectively. n denotes the sample size of the study.

Table A8.

Summary Statistics of Age (Mean±SD), BMI (Mean±SD), Sex [N(%)] and Minor Allele Frequencies of the Six Cohorts Stratified by Disease Status

Study	N	Age	BMI	Male	MAF rs6499640	MAF rs1121980
D2D2007	1699	59.5±8.3	27.3±5.0	729 (42.9)	0.417	0.402
DIAGEN	1058	61.0±14.1	28.1±5.5	458 (43.3)	0.398	0.466
FUSION S2	2219	59.3±8.2	28.7±5.1	1291 (58.2)	0.423	0.409
HUNT	1331	67.1±13.1	28.0±4.4	693 (52.1)	0.372	0.473
METSIM	1899	58.0±6.9	28.6±5.0	1899 (100)	0.411	0.444
TROMSO	1418	59.9±12.5	27.6±4.7	706 (49.8)	0.376	0.487
Total	9624	60.5±10.6	28.1±5.0	5776 (60.0)	0.403	0.441
D2D2007	458	63.4±7.5	30.5±5.5	271 (59.2)	0.385	0.404
DIAGEN	434	65.8±11.7	30.1±6.2	216 (49.8)	0.404	0.477
FUSION S2	1034	59.7±8.7	30.9±5.4	582 (56.3)	0.418	0.435
HUNT	578	68.9±11.4	29.2±4.6	299 (51.7)	0.379	0.501
METSIM	1209	60.5±6.6	30.2±5.2	1209 (100)	0.403	0.446

Study	N	Age	BMI	Male	MAF rs6499640	MAF rs1121980
TROMSO	709	59.9±12.5	29.2±4.9	352 (49.6)	0.373	0.491
Cases	4422	62.2±10.1	30.1±5.3	2929 (66.2)	0.397	0.456
D2D2007	1241	58.1±8.2	26.1±4.3	458 (36.9)	0.429	0.402
DIAGEN	624	57.6±14.6	26.7±4.4	242 (38.8)	0.393	0.458
FUSION S2	1185	58.9±7.6	26.9±3.9	709 (59.8)	0.428	0.387
HUNT	753	65.8±14.2	27.1±4.0	394 (52.3)	0.367	0.452
METSIM	690	53.7±5.0	25.9±3.1	690 (100)	0.425	0.442
TROMSO	709	59.9±12.5	25.9±3.8	354 (49.9)	0.378	0.483
Controls	5202	59.0±10.9	26.4±4.0	2847 (54.7)	0.408	0.429

Table A9.

Study-Specific Estimates of Odds Ratios [95% CI] and Associated p-Values across Six Case-Control Studies of Type 2 Diabetes

Study	$\exp(\widehat{\alpha}_G)$	p_{MA}	$\exp(\widehat{\beta}_G)$	p_G	$\exp(\widehat{\beta}_{CC})$	p_{CC}	$\exp(\widehat{\beta}_{EB})$	p_{EB}	p_{JOINT}	p_{MA+CC}	p_{MA+EB}
rs6499640											
D2D2007	0.91 [0.76–1.09]	0.301	0.92 [0.76–1.10]	0.347	1.05 [1.01–1.09]	*0.017*	1.04 [1.00–1.09]	0.065	*0.032*	*0.033*	0.106
DIAGEN	1.11 [0.92–1.36]	0.277	1.12 [0.92–1.36]	0.269	1.01 [0.97–1.05]	0.747	1.01 [0.98–1.05]	0.466	0.526	0.526	0.425
FUSION S2	0.95 [0.83–1.08]	0.431	0.92 [0.81–1.05]	0.220	1.06 [1.02–1.09]	*0.0006*	1.05 [1.02–1.08]	*0.003*	*0.002*	*0.002*	*0.010*
HUNT	1.07 [0.91–1.26]	0.394	1.07 [0.91–1.26]	0.395	1.00 [0.96–1.04]	0.954	1.01 [0.97–1.05]	0.731	0.694	0.694	0.655
METSIM	0.89 [0.74–1.06]	0.183	0.88 [0.74–1.07]	0.178	0.99 [0.94–1.05]	0.795	0.99 [0.95–1.03]	0.539	0.381	0.381	0.341
TROMSO	0.98 [0.83–1.15]	0.779	0.98 [0.83–1.15]	0.778	1.00 [0.96–1.04]	0.961	1.00 [0.87–1.04]	0.978	0.960	0.960	0.961
rs1121980											
D2D2007	0.97 [0.81–1.17]	0.776	0.98 [0.82–1.17]	0.808	0.98 [0.94–1.02]	0.286	0.99 [0.95–1.03]	0.564	0.548	0.543	0.813
DIAGEN	1.06 [0.88–1.29]	0.536	1.07 [0.88–1.30]	0.486	1.02 [0.97–1.06]	0.445	1.01 [0.97–1.04]	0.636	0.617	0.617	0.738
FUSION S2	1.19 [1.04–1.35]	*0.011*	1.19 [1.04–1.36]	*0.010*	0.99 [0.96–1.03]	0.710	1.00 [0.97–1.02]	0.761	*0.037*	*0.037*	*0.038*
HUNT	1.21 [1.03–1.41]	*0.020*	1.21 [1.03–1.41]	*0.020*	1.01 [0.97–1.05]	0.643	1.00 [0.97–1.04]	0.812	0.059	0.059	0.064

Study	$\exp(\widehat{\alpha}_G)$	p_{MA}	$\exp(\widehat{\beta}_G)$	p_G	$\exp(\widehat{\beta}_{CC})$	p_{CC}	$\exp(\widehat{\beta}_{EB})$	p_{EB}	p_{JOINT}	p_{MA+CC}	p_{MA+EB}
METSIM	0.98 [0.82–1.17]	0.839	0.98 [0.82–1.18]	0.838	1.00 [0.95–1.05]	0.968	1.01 [0.97–1.04]	0.719	0.979	0.979	0.918
TROMSO	0.98 [0.84–1.15]	0.822	0.98 [0.84–1.15]	0.819	1.02 [0.98–1.06]	0.399	1.02 [0.99–1.06]	0.188	0.683	0.684	0.411

The marginal SNP effect $\widehat{\alpha}_G$ was obtained by fitting Model 1; $\widehat{\beta}_G$, $\widehat{\beta}_{CC}$, and $\widehat{\beta}_{EB}$ were obtained by fitting Model 2.

p_{MA}, p_G, p_{CC}, and p_{EB} denote p-values associated with their corresponding estimated coefficients. P-values less than 0.05 are italicized.

Table A10.

Genome-wide Significant SNPs in North Finland Birth Cohort 1966 Study

GEMMA		ASSET		pASTA(MA)		pASTA(JOINT)	
SNP	p-values	SNP	p-values	SNP	p-values	SNP	p-values
rs3764261	5.5×10^{-30}	rs1811472	1.1×10^{-11}	rs3764261	3.6×10^{-20}	rs3764261	7.1×10^{-20}
rs1532624	2.8×10^{-24}	rs11265260	3.5×10^{-11}	rs1532624	1.7×10^{-17}	rs1532624	2.5×10^{-16}
rs7499892	2.2×10^{-17}	rs560887	8.2×10^{-11}	rs646776	5.2×10^{-14}	rs646776	4.3×10^{-13}
rs646776	8.1×10^{-16}	rs9989419	1.5×10^{-10}	rs7499892	4.1×10^{-12}	rs7499892	1.3×10^{-11}
rs2794520	9.4×10^{-16}	rs2650000	1.8×10^{-10}	rs2794520	3.0×10^{-11}	rs2794520	4.2×10^{-11}
rs1532085	8.7×10^{-13}	rs754524	1.5×10^{-9}	rs6728178	1.6×10^{-10}	rs6728178	1.0×10^{-9}
rs6989593	1.3×10^{-11}	rs12093699	2.9×10^{-8}	rs6754295	2.9×10^{-10}	rs6754295	1.8×10^{-9}
rs2592887	1.6×10^{-11}			rs693	9.5×10^{-10}	rs693	5.0×10^{-9}
rs1811472	7.6×10^{-10}			rs754524	9.9×10^{-10}	rs3923037	6.3×10^{-9}
rs7953249	9.9×10^{-10}			rs673548	1.0×10^{-9}	rs673548	6.6×10^{-9}
rs11265260	4.3×10^{-9}			rs676210	1.5×10^{-9}	rs754524	6.7×10^{-9}
rs2650000	1.2×10^{-8}			rs3923037	2.2×10^{-9}	rs676210	1.0×10^{-8}
rs754524	1.2×10^{-8}			rs207150	1.2×10^{-8}	rs207150	3.4×10^{-8}
rs174546	1.6×10^{-8}			rs157580	1.4×10^{-8}		
rs9989419	1.7×10^{-8}			rs2075650	2.2×10^{-8}		
rs560887	1.8×10^{-8}			rs754523	2.6×10^{-8}		
rs102275	1.9×10^{-8}			rs1429974	2.8×10^{-8}		
rs12093699	2.0×10^{-8}						
rs174556	3.0×10^{-8}						
rs2075650	4.0×10^{-8}						
rs415799	4.4×10^{-8}						
rs174537	4.4×10^{-8}						
rs2072114	5.0×10^{-8}						

The genome-wide significance level is 5.0×10^{-8}.

Table A11.

Combined p-Values for the Top 100 SNPs that Showed the Strongest Marginal Associations Using GEMMA

SNP	pASTA (MA)	Associated phenotypes	pASTA (JOINT)	Associated phenotypes	ASSET[†]	Associated phenotypes[†]
rs207150	1.2×10^{-8}	TC, LDL	3.4×10^{-8}	TC, LDL		
rs1892534	3.8×10^{-5}	CRP	8.1×10^{-5}	CRP, Glucose	2.0×10^{-7}	CRP, HDL, DBP
rs2211651	5.3×10^{-5}	CRP	1.2×10^{-4}	CRP, Glucose	3.3×10^{-7}	CRP, HDL, DBP
rs2889195	5.6×10^{-5}	CRP	1.2×10^{-4}	CRP, Glucose	4.8×10^{-7}	CRP, HDL, DBP
rs4420065	4.2×10^{-5}	CRP	1.0×10^{-4}	CRP, Glucose	2.3×10^{-7}	CRP, HDL, DBP
rs12753193	6.6×10^{-5}	CRP	1.4×10^{-4}	CRP, Glucose	4.0×10^{-7}	CRP, HDL, DBP
rs3738780	3.7×10^{-4}	Glucose, TC, LDL	5.8×10^{-4}	TC, LDL	3.8×10^{-4}	CRP, Glucose, TC, HDL, LDL, SBP
rs611917	2.2×10^{-7}	TC, LDL	5.5×10^{-7}	TC, LDL	2.4×10^{-7}	Glucose, HDL, LDL, DBP
rs646776	5.2×10^{-14}	TC, LDL	4.3×10^{-13}	TC, LDL		
rs2629824	4.4×10^{-2}	TC, HDL, LDL, TG	8.1×10^{-2}	TC, HDL, TG	4.8×10^{-2}	TC, HDL, TG
rs1811472	7.1×10^{-7}	CRP	1.6×10^{-7}	CRP	9.7×10^{-12}	CRP, TC, HDL, SBP
rs12093699	1.2×10^{-5}	CRP	1.6×10^{-5}	CRP	1.9×10^{-8}	CRP, TG, SBP
rs2592887	5.6×10^{-8}	CRP	1.3×10^{-7}	CRP		
rs2794520	3.0×10^{-11}	CRP	4.2×10^{-11}	CRP		
rs11265260	8.7×10^{-7}	CRP	2.1×10^{-6}	CRP	2.7×10^{-11}	CRP, TC, HDL
rs4844614	4.3×10^{-5}	TC, LDL, DBP	1.4×10^{-4}	TC, LDL	2.9×10^{-4}	CRP, Glucose, Insulin, TC, HDL, LDL, DBP
rs11894335	7.6×10^{-3}	HDL	7.6×10^{-4}	Glucose, Insulin, HDL, SBP	3.5×10^{-3}	Insulin, HDL, SBP
rs2881925	8.9×10^{-4}	Insulin, TC, HDL, LDL	1.4×10^{-3}	Insulin, TC, HDL, LDL, SBP	4.8×10^{-4}	Glucose, Insulin, HDL, LDL
rs3923037	2.2×10^{-9}	TC, HDL, LDL, TG	6.3×10^{-9}	TC, HDL, LDL, TG		
rs6728178	1.7×10^{-10}	TC, HDL, LDL, TG	1.0×10^{-9}	TC, HDL, LDL, TG		
rs6754295	2.9×10^{-10}	TC, HDL, LDL, TG	1.8×10^{-9}	TC, HDL, LDL, TG		
rs676210	2.9×10^{-9}	TC, HDL, LDL, TG	1.0×10^{-8}	TC, HDL, LDL, TG		
rs693	9.5×10^{-10}	TC, LDL, TG	5.0×10^{-9}	TC, LDL		
rs673548	1.0×10^{-9}	TC, HDL, LDL, TG	6.6×10^{-9}	TC, HDL, LDL, TG		
rs1429974	2.7×10^{-8}	TC, LDL	1.8×10^{-7}	TC, LDL		

SNP	pASTA (MA)	Associated phenotypes	pASTA (JOINT)	Associated phenotypes	ASSET[†]	Associated phenotypes[†]
rs754524	9.9×10^{-10}	TC, LDL	6.7×10^{-9}	TC, LDL	1.5×10^{-9}	Glucose, HDL, LDL
rs754523	2.6×10^{-8}	TC, LDL	1.7×10^{-7}	TC, LDL		
rs6711889	9.8×10^{-4}	Glucose, Insulin, HDL, TG	2.5×10^{-3}	CRP, Glucose, Insulin, HDL, LDL, TG	5.5×10^{-5}	Glucose, Insulin, HDL, TG
rs1260326	1.3×10^{-5}	TG	3.4×10^{-5}	TG		
rs780094	3.3×10^{-5}	TG	9.1×10^{-5}	TG		
rs560887	1.9×10^{-6}	Glucose	3.1×10^{-6}	Glucose	8.2×10^{-11}	CRP, Glucose, Insulin, HDL, TG
rs563694	1.1×10^{-4}	Glucose	1.6×10^{-4}	Glucose	4.4×10^{-7}	CRP, Glucose, TC
rs1513342	5.3×10^{-3}	Glucose, LDL, SBP	1.1×10^{-2}	Glucose, TC, HDL, LDL, SBP	1.6×10^{-3}	CRP, Glucose, Insulin, LDL, SBP, DBP
rs10513660	1.3×10^{-3}	CRP, LDL, SBP, DBP	7.4×10^{-3}	CRP, LDL, SBP, DBP	1.0×10^{-3}	CRP, Insulin, LDL, SBP
rs2962663	5.0×10^{-2}	Glucose, LDL, TG, DBP	5.2×10^{-2}	Glucose, Insulin, HDL, LDL, TG, DBP	2.3×10^{-2}	CRP, Glucose, Insulin, HDL, LDL, TG, DBP
rs2840796	3.0×10^{-3}	Glucose, Insulin, HDL, LDL	1.4×10^{-2}	Glucose	7.2×10^{-5}	Glucose, Insulin, HDL, LDL
rs2171981	4.6×10^{-6}	TC, HDL, LDL	1.0×10^{-5}	Glucose, TC, HDL, LDL		
rs10244051	6.0×10^{-5}	Glucose	1.4×10^{-4}	Glucose	4.9×10^{-7}	Glucose, SBP, DBP
rs2191348	7.4×10^{-5}	Glucose	1.7×10^{-4}	Glucose	3.3×10^{-7}	Glucose, SBP, DBP
rs2971671	1.2×10^{-4}	Glucose	1.8×10^{-4}	Glucose	2.6×10^{-7}	Glucose, Insulin, HDL, LDL, SBP, DBP
rs2908290	2.3×10^{-4}	Glucose, TG	2.3×10^{-4}	Glucose, Insulin, TG	9.5×10^{-6}	Glucose, Insulin, HDL, TG, SBP
rs3801975	3.8×10^{-5}	Insulin, TC, LDL, SBP, DBP	8.5×10^{-5}	Glucose, Insulin, TC, LDL, DBP		
rs983309	2.6×10^{-7}	CRP, TC, HDL, LDL	1.3×10^{-6}	CRP, TC, HDL, LDL		
rs7002551	3.4×10^{-5}	CRP, Glucose, TC, HDL, LDL	2.3×10^{-4}	CRP, Glucose, TC, HDL, LDL		
rs12543276	2.5×10^{-5}	CRP, TC, HDL, LDL	1.5×10^{-4}	CRP, TC, HDL, LDL		
rs2126259	6.0×10^{-8}	CRP, Glucose, TC, HDL, LDL	3.3×10^{-7}	CRP, TC, HDL, LDL, DBP		
rs7004769	2.1×10^{-5}	CRP, TC, HDL, LDL	1.3×10^{-4}	CRP, TC, HDL, LDL, SBP, DBP		
rs10096633	7.9×10^{-8}	HDL, TG	2.0×10^{-7}	HDL, TG		
rs4523270	2.0×10^{-4}	HDL, TG	1.3×10^{-5}	CRP, TC, HDL, LDL, TG	1.6×10^{-5}	HDL, TG

SNP	pASTA (MA)	Associated phenotypes	pASTA (JOINT)	Associated phenotypes	ASSET[†]	Associated phenotypes[†]
rs2083637	2.7×10^{-5}	HDL, TG	2.5×10^{-6}	CRP, TC, HDL, LDL, TG	3.6×10^{-6}	HDL, TG
rs6989593	3.8×10^{-1}	LDL, DBP	3.8×10^{-1}	TC, HDL, LDL, SBP, DBP	3.0×10^{-1}	Glucose, Insulin, TC, LDL, SBP, DBP
rs2150702	1.7×10^{-3}	CRP, TG, DBP	5.4×10^{-3}	CRP, TG, SBP, DBP	1.9×10^{-4}	CRP, Glucose, Insulin, HDL, LDL, TG, DBP
rs2807580	2.9×10^{-3}	CRP, Insulin, TC, HDL, TG, SBP, DBP	1.2×10^{-2}	CRP, Insulin, TC, HDL, TG, SBP	5.5×10^{-4}	CRP, Insulin, HDL, TG, SBP
rs9411438	8.1×10^{-3}	HDL, TG, SBP, DBP	8.4×10^{-3}	CRP, HDL, TG, SBP, DBP	4.5×10^{-3}	Glucose, HDL, LDL, SBP
rs943451	2.3×10^{-3}	Glucose, Insulin, TC, HDL, LDL, SBP, DBP	4.1×10^{-3}	TC, HDL, LDL, SBP, DBP	8.7×10^{-3}	Glucose, Insulin, HDL, LDL, TG, SBP, DBP
rs174537	1.7×10^{-6}	Glucose, TC, LDL	1.3×10^{-5}	Glucose, TC, LDL		
rs102275	9.1×10^{-7}	Glucose, TC, HDL, LDL	8.2×10^{-6}	Glucose, TC, LDL		
rs174546	9.7×10^{-7}	Glucose, TC, LDL	8.2×10^{-6}	Glucose, TC, LDL		
rs174556	1.0×10^{-6}	Glucose, TC, LDL	1.0×10^{-5}	Glucose, TC, LDL		
rs174570	8.0×10^{-7}	Glucose, TC, LDL	4.2×10^{-6}	Glucose, Insulin, TC, LDL		
rs1535	1.7×10^{-6}	Glucose, TC, LDL	1.3×10^{-5}	Glucose, TC, LDL		
rs2072114	1.6×10^{-7}	Glucose, TC, LDL	1.7×10^{-6}	Glucose, TC, LDL		
rs2851682	2.0×10^{-7}	Glucose, TC, LDL	1.3×10^{-6}	Glucose, Insulin, TC, LDL		
rs3847554	1.7×10^{-4}	Glucose	4.3×10^{-4}	Glucose	2.9×10^{-6}	CRP, Glucose, HDL, LDL, TG
rs1447352	2.4×10^{-4}	Glucose	2.1×10^{-4}	Glucose, LDL	6.2×10^{-7}	CRP, Glucose, LDL
rs7121092	2.0×10^{-4}	Glucose	1.7×10^{-4}	Glucose, LDL	3.9×10^{-7}	CRP, Glucose, LDL
rs598693	2.9×10^{-2}	Glucose, Insulin, TC, LDL, DBP	2.6×10^{-1}	Glucose, Insulin, TC, DBP	1.9×10^{-2}	Glucose, Insulin, TC, HDL, DBP
rs532107	3.0×10^{-3}	Glucose, Insulin, HDL, LDL, TG, DBP	3.9×10^{-3}	Insulin, TC, HDL, LDL, TG, SBP, DBP		
rs541407	2.2×10^{-4}	CRP, Glucose, Insulin, HDL, LDL, TG, SBP, DBP	6.0×10^{-4}	Glucose, Insulin, TC, HDL, LDL, TG, SBP, DBP		
rs2000571	1.3×10^{-4}	CRP, TC, HDL, LDL, TG	3.9×10^{-4}	Insulin, TC, HDL, LDL, TG	9.8×10^{-6}	CRP, Glucose, Insulin, TC, HDL, TG

SNP	pASTA (MA)	Associated phenotypes	pASTA (JOINT)	Associated phenotypes	ASSET[†]	Associated phenotypes[†]
rs2650000	6.2×10^{-7}	CRP	1.1×10^{-6}	CRP	1.3×10^{-10}	CRP, TC, HDL
rs7953249	3.0×10^{-7}	CRP	6.2×10^{-7}	CRP	6.3×10^{-12}	HDL, CRP
rs1169300	7.1×10^{-5}	CRP	1.4×10^{-4}	CRP	8.9×10^{-8}	CRP, HDL, LDL, TG
rs2464196	7.0×10^{-5}	CRP	1.5×10^{-4}	CRP	8.8×10^{-8}	CRP, HDL, LDL, TG
rs735396	7.3×10^{-5}	CRP	1.7×10^{-4}	CRP	7.3×10^{-8}	CRP, HDL, LDL, TG
rs2806639	6.0×10^{-3}	TG, SBP, DBP	3.5×10^{-2}	TG, SBP	9.2×10^{-4}	Insulin, HDL, LDL, TG, SBP
rs1570107	1.1×10^{-2}	CRP, HDL, DBP	1.2×10^{-2}	CRP, HDL, DBP	6.0×10^{-3}	CRP, HDL, TG, DBP
rs10131303	1.2×10^{-2}	CRP, HDL, DBP	1.4×10^{-2}	CRP, HDL, DBP	5.4×10^{-3}	CRP, HDL, TG, DBP
rs166358	5.6×10^{-4}	HDL	1.4×10^{-3}	HDL		
rs1532085	3.2×10^{-7}	TC, HDL	9.0×10^{-7}	TC, HDL		
rs415799	1.0×10^{-4}	TC, HDL	1.7×10^{-4}	TC, HDL		
rs473224	1.1×10^{-4}	TC, HDL, TG	7.1×10^{-4}	TC, HDL, TG		
rs261336	5.1×10^{-5}	TC, HDL, TG	3.7×10^{-4}	TC, HDL, TG		
rs9989419	1.0×10^{-6}	HDL	2.7×10^{-6}	HDL	1.3×10^{-10}	CRP, Glucose, HDL
rs3764261	3.5×10^{-20}	HDL	7.1×10^{-20}	HDL		
rs1532624	1.7×10^{-16}	HDL	2.5×10^{-16}	HDL		
rs7499892	4.1×10^{-12}	HDL	1.3×10^{-11}	HDL		
rs2271308	1.3×10^{-4}	TC, HDL, LDL	7.4×10^{-4}	CRP, Insulin, TC, HDL, LDL		
rs1482275	2.0×10^{-4}	Insulin, TC, LDL, TG	1.5×10^{-3}	Insulin, TC, LDL	7.4×10^{-4}	Insulin, HDL, LDL, TG
rs11082819	3.2×10^{-2}	CRP, Glucose, LDL, TG, SBP	6.4×10^{-2}	Glucose, TG, SBP	1.3×10^{-2}	CRP, Glucose, Insulin, HDL, LDL, TG, SBP
rs2000731	2.7×10^{-2}	CRP, Glucose, LDL, TG, SBP	5.5×10^{-2}	Glucose, TG, SBP	1.2×10^{-2}	CRP, Glucose, Insulin, HDL, LDL, TG, SBP
rs11668477	7.7×10^{-7}	TC, LDL	3.7×10^{-6}	TC, LDL		
rs2228671	2.2×10^{-7}	TC, LDL	1.4×10^{-6}	TC, LDL		
rs7256832	1.9×10^{-4}	Insulin, TC, LDL, DBP	1.0×10^{-3}	TC, LDL, DBP	9.0×10^{-5}	Insulin, LDL, DBP
rs157580	1.4×10^{-8}	TC, HDL, LDL, TG	2.1×10^{-7}	CRP, TC, HDL, LDL, TG	7.2×10^{-8}	CRP, TC, HDL, LDL, TG
rs2075650	2.2×10^{-8}	CRP, TC, LDL, TG	3.5×10^{-7}	CRP, TC, LDL, TG	7.1×10^{-10}	CRP, HDL, LDL, TG
rs405509	7.0×10^{-4}	CRP, LDL	1.4×10^{-3}	CRP, LDL	3.6×10^{-5}	CRP, LDL
rs3786625	5.6×10^{-3}	Glucose, SBP, DBP	1.9×10^{-2}	SBP	1.2×10^{-3}	Insulin, HDL, TG, SBP
rs1800961	3.5×10^{-4}	HDL	5.5×10^{-4}	HDL	5.0×10^{-6}	CRP, HDL, TG, DBP

SNP	pASTA (MA)	Associated phenotypes	pASTA (JOINT)	Associated phenotypes	ASSET[†]	Associated phenotypes[†]
rs2837324	5.3×10^{-2}	Glucose, HDL, LDL	1.1×10^{-1}	CRP, Glucose, Insulin, HDL, LDL	6.7×10^{-2}	Glucose, HDL, LDL

[†] Some of the cells in these two columns are left blank due to computational issue in the ASSET package.

Abbreviations: CRP, C-reactive Protein; TC, Total Cholesterol; HDL, High Density Lipoprotein; LDL, Low Density Lipoprotein; TG, Triglycerides; Systolic Blood Pressure (SBP); Diastolic Blood Pressure (DBP).

Table A12.

Estimates of Linear Regression Coefficients [95% CI] and p-values Corresponding to the Nine Phenotypes in the NFBC 1966 Study

Phenotype	$\hat{\alpha}_G$	p_{MA}	$\hat{\beta}_G$	p_G	$\hat{\beta}_{GE}$	p_{GE}	p_{JOINT}
rs2083637							
CRP	0.02 [−0.02, 0.06]	3.5×10^{-1}	0.02 [−0.02, 0.06]	3.5×10^{-1}	−0.07 [−0.11, −0.03]	1.3×10^{-3}	3.6×10^{-3}
Glucose	−0.01 [−0.05, 0.04]	7.9×10^{-1}	−0.01 [−0.05, 0.04]	7.9×10^{-1}	0.004 [−0.05, 0.04]	8.7×10^{-1}	9.5×10^{-1}
Insulin	0.03 [−0.01, 0.08]	1.6×10^{-1}	0.03 [−0.01, 0.08]	1.6×10^{-1}	0.01 [−0.04, 0.05]	7.0×10^{-1}	3.4×10^{-1}
TC	−0.04 [−0.08, 0.01]	1.2×10^{-1}	−0.04 [−0.08, 0.01]	1.1×10^{-1}	−0.06 [−0.11, −0.02]	6.3×10^{-3}	6.8×10^{-3}
HDL	−0.10 [−0.15, −0.06]	3.4×10^{-6}	−0.10 [−0.15, −0.06]	3.4×10^{-6}	−0.003 [−0.05, 0.04]	9.0×10^{-1}	2.0×10^{-5}
LDL	−0.02 [−0.07, 0.02]	3.1×10^{-1}	−0.02 [−0.07, 0.02]	3.1×10^{-1}	−0.06 [−0.10, −0.02]	8.2×10^{-3}	1.8×10^{-2}
TG	0.10 [0.06, 0.14]	5.6×10^{-6}	0.10 [0.06, 0.14]	5.7×10^{-6}	−0.03 [−0.08, 0.01]	1.1×10^{-1}	8.9×10^{-6}
SBP	−0.03 [−0.08, 0.01]	1.6×10^{-1}	−0.03 [−0.08, 0.01]	1.6×10^{-1}	−0.01 [−0.05, 0.04]	7.3×10^{-1}	3.5×10^{-1}
DBP	−0.02 [−0.06, 0.03]	4.2×10^{-1}	−0.02 [−0.06, 0.03]	4.2×10^{-1}	0.02 [−0.03, 0.07]	3.9×10^{-1}	5.0×10^{-1}
rs754524							
CRP	0.003 [−0.04, 0.04]	8.8×10^{-1}	0.003 [−0.04, 0.04]	8.7×10^{-1}	0.01 [−0.03, 0.05]	5.8×10^{-1}	8.5×10^{-1}
Glucose	0.03 [−0.01, 0.07]	1.8×10^{-1}	0.03 [−0.01, 0.07]	1.8×10^{-1}	−0.005 [−0.05, 0.04]	8.3×10^{-1}	4.0×10^{-1}
Insulin	0.02 [−0.03, 0.06]	4.7×10^{-1}	0.02 [−0.03, 0.06]	4.6×10^{-1}	0.01 [−0.03, 0.05]	6.8×10^{-1}	7.0×10^{-1}
TC	−0.13 [−0.18, −0.09]	6.3×10^{-10}	−0.13 [−0.18, −0.09]	6.0×10^{-10}	−0.01 [−0.05, 0.03]	6.4×10^{-1}	4.3×10^{-9}
HDL	0.04 [0.003, 0.09]	3.6×10^{-2}	0.04 [0.003, 0.09]	3.5×10^{-2}	0.02 [−0.02, 0.06]	4.1×10^{-1}	7.9×10^{-2}
LDL	−0.14 [−0.19, −0.10]	2.1×10^{-11}	−0.14 [−0.19, −0.10]	2.0×10^{-11}	−0.01 [−0.05, 0.03]	6.8×10^{-1}	1.5×10^{-10}
TG	−0.05 [−0.09, −0.01]	1.8×10^{-2}	−0.05 [−0.09, −0.01]	1.8×10^{-2}	−0.004 [−0.05, 0.04]	8.5×10^{-1}	5.9×10^{-2}
SBP	0.01 [−0.03, 0.06]	5.4×10^{-1}	0.01 [−0.03, 0.06]	5.2×10^{-1}	0.03 [−0.01, 0.07]	1.9×10^{-1}	3.5×10^{-1}

Phenotype	$\widehat{\alpha}_G$	p_{MA}	$\widehat{\beta}_G$	p_G	$\widehat{\beta}_{GE}$	p_{GE}	p_{JOINT}
DBP	0.01 [−0.04, 0.05]	7.2×10^{-1}	0.01 [−0.03, 0.05]	7.0×10^{-1}	0.03 [−0.01, 0.07]	1.7×10^{-1}	3.6×10^{-1}

Abbreviations: CRP, C-reactive Protein; TC, Total Cholesterol; HDL, High Density Lipoprotein; LDL, Low Density Lipoprotein; TG, Triglycerides; Systolic Blood Pressure (SBP); Diastolic Blood Pressure (DBP).

The marginal SNP effect $\widehat{\alpha}_G$ is obtained by fitting Model 1; $\widehat{\beta}_G$ and $\widehat{\beta}_{GE}$ are obtained by fitting Model 2. p_{MA}, p_G, and p_{GE} denote the p-values associated with their corresponding estimated coefficients. P-values less than 0.05 are italicized.

References

1. Visscher PM, Brown MA, McCarthy MI, Yang J: Five Years of GWAS Discovery. Am J Hum Genet 2012;90:7–24. [PubMed: 22243964]
2. Willer CJ, Li Y, Abecasis GR: METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191. [PubMed: 20616382]
3. Lin DY, Zeng D: Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. Genet Epidemiol 2010;34:60–66. [PubMed: 19847795]
4. Han B, Eskin E: Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. Am J Hum Genet 2011;88:586–598. [PubMed: 21565292]
5. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al.: Pervasive Sharing of Genetic Effects in Autoimmune Disease. PLOS Genet 2011;7:e1002254. [PubMed: 21852963]
6. Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al.: Meta-analysis of Correlated Traits via Summary Statistics from GWASs with an Application in Hypertension. Am J Hum Genet 2015;96:21–36. [PubMed: 25500260]
7. Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, et al.: A Subset-Based Approach Improves Power and Interpretation for the Combined Analysis of Genetic Association Studies of Heterogeneous Traits. Am J Hum Genet 2012;90:821–835. [PubMed: 22560090]
8. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al.: Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 2016;48:510–518. [PubMed: 26974007]
9. Hung RJ, Ulrich CM, Goode EL, Brhane Y, Muir K, Chan AT, et al.: Cross Cancer Genomic Investigation of Inflammation Pathway for Five Common Cancers: Lung, Ovary, Prostate, Breast, and Colorectal Cancer. JNCI J Natl Cancer Inst 2015;107 DOI: 10.1093/jnci/djv246
10. Boonstra PS, Mukherjee B, Gruber SB, Ahn J, Schmit SL, Chatterjee N: Tests for Gene-Environment Interactions and Joint Effects With Exposure Misclassification. Am J Epidemiol 2016;183:237–247. [PubMed: 26755675]
11. Dai JY, Logsdon BA, Huang Y, Hsu L, Reiner AP, Prentice RL, et al.: Simultaneously testing for marginal genetic association and gene-environment interaction. Am J Epidemiol 2012;176:164–173. [PubMed: 22771729]
12. Kraft P, Yen Y-C, Stram DO, Morrison J, Gauderman WJ: Exploiting Gene-Environment Interaction to Detect Genetic Associations. Hum Hered 2007;63:111–119. [PubMed: 17283440]
13. Thomas D: Gene–environment-wide association studies: emerging approaches. Nat Rev Genet 2010;11:259–272. [PubMed: 20212493]
14. Piegorsch WW, Weinberg CR, Taylor JA: Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. Stat Med 1994;13:153–162. [PubMed: 8122051]
15. Mukherjee B, Chatterjee N: Exploiting Gene-Environment Independence for Analysis of Case-Control Studies: An Empirical Bayes-Type Shrinkage Estimator to Trade-Off between Bias and Efficiency. Biometrics 2008;64:685–694. [PubMed: 18162111]
16. Manning AK, LaValley M, Liu C-T, Rice K, An P, Liu Y, et al.: Meta-analysis of Gene-Environment interaction: joint estimation of SNP and SNP×Environment regression coefficients. Genet Epidemiol 2011;35:11–18. [PubMed: 21181894]

17. Li S, Mukherjee B, Taylor JMG, Rice KM, Wen X, Rice JD, et al.: The role of environmental heterogeneity in meta-analysis of gene-environment interactions with quantitative traits. Genet Epidemiol 2014;38:416–429. [PubMed: 24801060]

18. Aschard H, Hancock DB, London SJ, Kraft P: Genome-Wide Meta-Analysis of Joint Tests for Genetic and Gene-Environment Interaction Effects. Hum Hered 2011;70:292–300.

19. Fisher RA: Statistical Methods For Research Workers. Cosmo Publications, 1925.

20. Dai JY, Kooperberg C, Leblanc M, Prentice RL: Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. Biometrika 2012;99:929–944. [PubMed: 23843674]

21. Taylor JE, Worsley KJ, Gosselin F: Maxima of discretely sampled random fields, with an application to 'bubbles.' Biometrika 2007;94:1–18.

22. Lin DY, Sullivan PF: Meta-analysis of genome-wide association studies with overlapping subjects. Am J Hum Genet 2009;85:862–872. [PubMed: 20004761]

23. Yang J, Ferreira T, Morris AP, Medland SE, Consortium GI of AnT (GIANT), Consortium DiaGRAM (DIAGRAM), et al.: Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 2012;44:369–375. [PubMed: 22426310]

24. Yang JJ: Distribution of Fisher's combination statistic when the tests are dependent. J Stat Comput Simul 2010;80:1–12.

25. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al.: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341–1345. [PubMed: 17463248]

26. Estes JP, Rice JD, Li S, Stringham HM, Boehnke M, Mukherjee B: Meta-analysis of gene-environment interaction exploiting gene-environment independence across multiple case-control studies. Stat Med 2017; DOI: 10.1002/sim.7398

27. Sabatti C, Service SK, Hartikainen A-L, Pouta A, Ripatti S, Brodsky J, et al.: Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet 2009;41:35–46. [PubMed: 19060910]

28. Zhou X, Stephens M: Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods 2014;11:407–409. [PubMed: 24531419]

29. Zhou X: A unified framework for variance component estimation with summary statistics in genome-wide association studies. Ann Appl Stat 2017;11:2027–2051. [PubMed: 29515717]

30. Zhou X, Stephens M: Genome-wide efficient mixed-model analysis for association studies. Nat Genet 2012;44:821–824. [PubMed: 22706312]

31. Bhattacharjee S, Chatterjee N, Han S, Wheeler W: CGEN: An R package for analysis of case-control studies in genetic… in CGEN: An R package for analysis of case-control studies in genetic epidemiology 2012 [cited 2017 Aug 9];Available from: https://rdrr.io/bioc/CGEN/man/CGEN.html

32. Egaña-gorroño L, Martínez E, Cormand B, Escribà T, Gatell J, Arnedo M: Impact of genetic factors on dyslipidemia in Hiv-infected patients starting antiretroviral therapy. Aids 2013;27:529–538. [PubMed: 23262498]

33. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M: A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet 2012;44:1066–1071. [PubMed: 22902788]

34. Sun S, Hood M, Scott L, Peng Q, Mukherjee S, Tung J, et al.: Differential expression analysis for RNAseq using Poisson mixed models. Nucleic Acids Res 2017;45:e106. [PubMed: 28369632]

35. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al.: Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 2016;167:1853–1866.e17. [PubMed: 27984732]

36. Liu Z, Lin X: Multiple Phenotype Association Tests Using Summary Statistics in Genome-Wide Association Studies. Biometrics 2018;74:165–175. [PubMed: 28653391]

37. Dutta D, Scott L, Boehnke M, Lee S: Multi-SKAT: General framework to test multiple phenotype associations of rare variants. Genet Epidemiol In Press;

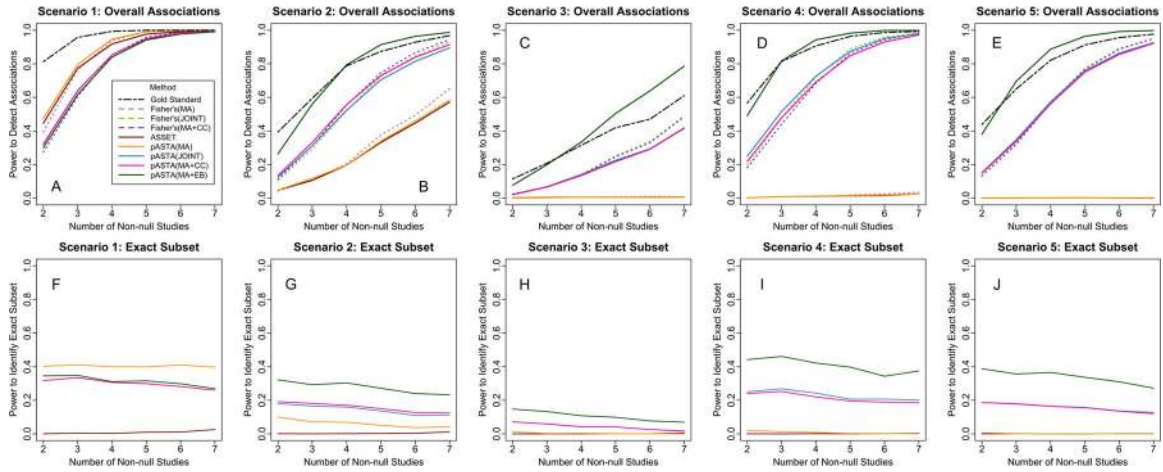38. Mersmann O: microbenchmark: Accurate Timing Functions. 2015.

Figure 1. Power Comparison of Nine Meta-Analysis Methods for Detecting Overall Associations and Identifying True Non-Null Studies in Case-Control Studies

Empirical power curves (top row) at significance level of 0.001 are shown for nine meta-analysis methods (Table 1). Power curves of identifying exact subset (bottom row) are plotted only for subset-based meta-analysis method ASSET and pASTA. In each simulation, a total of ten studies are analyzed, and each study contains 6,000 cases and 6,000 controls. Scenario 1 considers no environmental main effect or interaction, representing a situation where only marginal genetic effects exist (data generated from Model 1); Scenario 2 assumes strong interaction and moderate genetic main effect; Scenario 3 assumes strong interaction and no genetic main effect; Scenario 4 assumes positive interactions and negative genetic main effects, resulting in nearly-null marginal genetic effects; Scenario 5 considers negative interaction and positive genetic main effect.

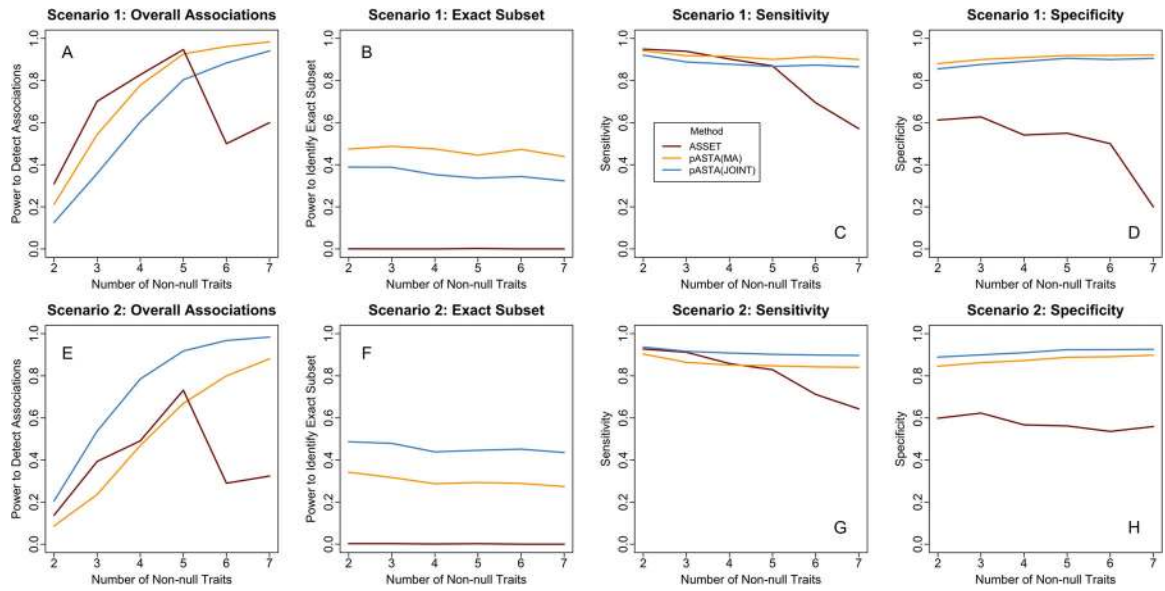**Figure 2. Power Comparison of Subset-Based Meta-Analysis Methods Applied to Correlated Quantitative Phenotypes**

A cohort study with nine correlated phenotypes measured and 6,000 subjects are simulated. The performance of the subset-based methods is evaluated in the absence (top row) and presence (bottom row) of gene-environment interaction. The level of significance is 0.001, and simulations are repeated 2,000 times.
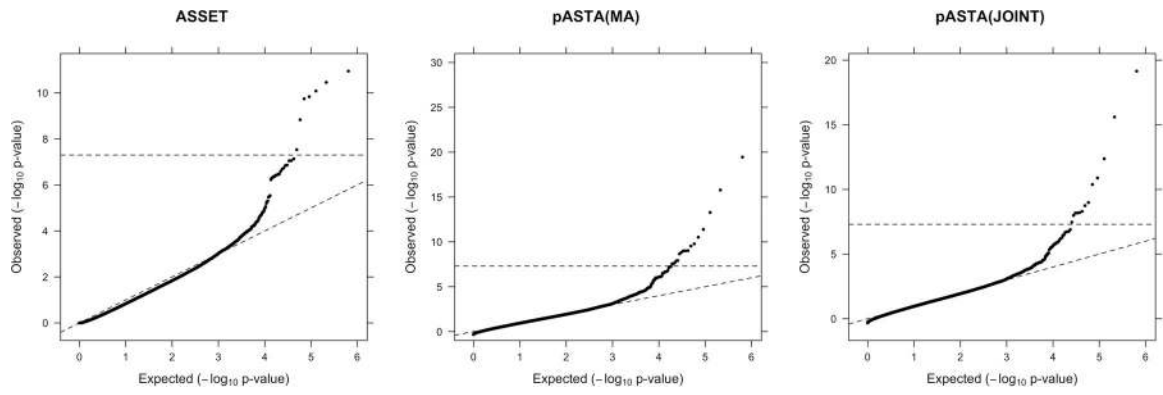
Figure 3. Quantile-Quantile Plots of p-Values Obtained Using ASSET, pASTA(MA), and pASTA(JOINT) in the Genome-Wide Analysis of the NFBC1966 Study.
The horizontal dashed line indicates the genome-wide level of significance 5×10^{-8}.

Table 1.

Methods Considered in Simulation Studies

Method Label	Explanation
Fisher's(MA)	Fisher's combined p-value method applied to p-values obtained from testing hypotheses of MA, i.e. $\alpha_G^{(k)} = 0$.
Fisher's(JOINT)	Fisher's combined p-value method applied to p-values obtained from testing hypotheses of JOINT, i.e. $\beta_G^{(k)} = \beta_{GE}^{(k)} = 0$.
Fisher's(MA+CC)	Fisher's combined p-value method applied to p-values obtained from testing hypotheses of MA+GE, i.e. $\alpha_G^{(k)} = \beta_{GE}^{(k)} = 0$; $\beta_{GE}^{(k)}$ being CC estimator.
ASSET(MA)	Two-sided ASSET for testing hypotheses of MA, i.e. $\alpha_G^{(k)} = 0$
pASTA(MA)	pASTA applied to p-values obtained from testing hypotheses of MA, i.e. $\alpha_G^{(k)} = 0$.
pASTA(JOINT)	pASTA applied to p-values obtained from testing hypotheses of JOINT, i.e. $\beta_G^{(k)} = \beta_{GE}^{(k)} = 0$.
pASTA(MA+CC)	pASTA applied to p-values obtained from testing hypotheses of MA+GE, i.e. $\alpha_G^{(k)} = \beta_{GE}^{(k)} = 0$; $\beta_{GE}^{(k)}$ being CC estimator.
pASTA(MA+EB)	pASTA applied to p-values obtained from testing hypotheses of MA+GE, i.e. $\alpha_G^{(k)} = \beta_{GE}^{(k)} = 0$; $\beta_{GE}^{(k)}$ being EB estimator.
Gold standard	Fisher's combined p-value method as if the true subset of non-null studies and the model from which data were generated were known a priori.

The null hypothesis in gold standard depends on specific situations. In the presence and absence of G-E interaction, JOINT and MA were tested, respectively. Therefore, the statistical power is not comparable across scenarios.

Table 2.

Type I Error Rates at Various Significance Levels for the Two Simulation Studies

Method	Scenario 1[†]			Scenario 2[†]		
	$\alpha = 10^{-2}$	$\alpha = 10^{-3}$	$\alpha = 10^{-5}$	$\alpha = 10^{-2}$	$\alpha = 10^{-3}$	$\alpha = 10^{-5}$
Simulation 1[‡]						
Fisher's(MA)	9.4×10^{-3}	1.3×10^{-3}	1.2×10^{-5}	1.0×10^{-2}	9.0×10^{-4}	9.0×10^{-6}
Fisher's(JOINT)	1.2×10^{-2}	1.1×10^{-3}	1.1×10^{-5}	9.4×10^{-3}	7.0×10^{-4}	1.1×10^{-5}
Fisher's(MA+CC)	1.2×10^{-2}	1.1×10^{-3}	1.1×10^{-5}	9.8×10^{-3}	8.0×10^{-4}	1.0×10^{-5}
ASSET	1.1×10^{-2}	1.2×10^{-3}	1.4×10^{-5}	1.1×10^{-1}	8.0×10^{-4}	1.1×10^{-5}
pASTA(MA)	9.8×10^{-3}	9.0×10^{-4}	1.1×10^{-5}	1.0×10^{-2}	8.0×10^{-4}	1.2×10^{-5}
pASTA(JOINT)	1.1×10^{-2}	9.0×10^{-4}	1.1×10^{-5}	9.0×10^{-3}	7.0×10^{-4}	8.6×10^{-6}
pASTA(MA+CC)	1.1×10^{-2}	9.0×10^{-4}	1.1×10^{-5}	9.4×10^{-3}	6.0×10^{-4}	7.6×10^{-6}
pASTA(MA+EB)	6.8×10^{-3}	6.0×10^{-4}	3.8×10^{-6}	5.0×10^{-3}	3.0×10^{-4}	4.2×10^{-6}
Simulation 2[‡]						
ASSET	5.0×10^{-3}	5.0×10^{-4}	8.0×10^{-6}	3.4×10^{-3}	2.9×10^{-4}	4.0×10^{-6}
pASTA(MA)	7.5×10^{-3}	6.6×10^{-4}	6.0×10^{-6}	6.1×10^{-3}	5.7×10^{-4}	6.0×10^{-6}
pASTA(JOINT)	7.1×10^{-3}	5.6×10^{-4}	1.2×10^{-5}	5.5×10^{-3}	4.5×10^{-4}	6.0×10^{-6}

[†]Scenario 1 and Scenario 2 correspond to situations where environmental effects are absent and present, respectively. Simulations were replicated 5×10^5 times. Type I error was estimated as proportion of null hypotheses rejected.

[‡]Simulation 1 considers $K = 10$ independent case-control studies and Simulation 2 considers $K = 9$ correlated phenotypes.

Table 3.

Meta-analysis of Six Case-control Studies of Type 2 Diabetes

Methods	rs6499640		rs1121980	
	p-Value	Associated Subset	p-Value	Associated Subset
Fixed-Effect	0.497		*0.029*	
Fisher's(MA)	0.413		0.080	
Fisher's(JOINT)	*0.023*		0.227	
Fisher's(MA+CC)	*0.025*		0.228	
ASSET	0.490	D2D2007, DIAGEN, FUSION S2, HUNT, METSIM	*0.031*	D2D2007, FUSION S2, HUNT, METSIM, TROMSO
pASTA(MA)	0.923	D2D2007, DIAGEN, METSIM	*0.024*	FUSION S2, HUNT
pASTA(JOINT)	*0.007*	D2D2007, FUSION S2	0.142	FUSION S2, HUNT
pASTA(MA+CC)	*0.008*	D2D2007, FUSION S2	0.144	FUSION S2, HUNT
pASTA(MA+EB)	0.083	D2D2007, FUSION S2	0.154	FUSION S2, HUNT

The six studies are independent with no overlapping subjects. P-values less than 0.05 are italicized.

Table 4.

Joint Analysis of Multiple Phenotypes in NFBC 1966 Study

	rs2083637		rs754524	
Methods	**p-Value**	**Associated Subset**	**p-Value**	**Associated Subset**
pASTA(MA)	2.7×10^{-5}	HDL, TG	9.9×10^{-10}	TC, LDL
pASTA(JOINT)	2.5×10^{-6}	CRP, TC, HDL, LDL, TG	6.7×10^{-9}	TC, LDL
ASSET	3.6×10^{-6}	HDL, TG	1.5×10^{-9}	Glucose, HDL, LDL

Abbreviations: CRP, C-reactive Protein; TC, Total Cholesterol; HDL, High Density Lipoprotein; LDL, Low Density Lipoprotein; TG, Triglycerides; Systolic Blood Pressure (SBP); Diastolic Blood Pressure (DBP).

Correlations among study-specific Z-statistics were approximated using phenotypic correlations.