Title
Determining Which Phenotypes Underlie a Pleiotropic Signal

Journal
Genetic Epidemiology, 40(5)

ISSN
0741-0395

Authors
Majumdar, Arunabha
Haldar, Tanushree
Witte, John S

Publication Date
2016-07-01

DOI
10.1002/gepi.21973

Peer reviewed

HHS Public Access

Author manuscript

Genet Epidemiol. Author manuscript; available in PMC 2016 July 01.

Published in final edited form as:

Genet Epidemiol. 2016 July ; 40(5): 366–381. doi:10.1002/gepi.21973.

Determining which phenotypes underlie a pleiotropic signal

Arunabha Majumdar, **Tanushree Haldar**, and **John S. Witte***

Department of Epidemiology and Biostatistics, University of California, San Francisco

Abstract

Discovering pleiotropic loci is important to understand the biological basis of seemingly distinct phenotypes. Most methods for assessing pleiotropy only test for the overall association between genetic variants and multiple phenotypes. To determine which specific traits are pleiotropic, we evaluate via simulation and application three different strategies. The first is model selection techniques based on the inverse regression of genotype on phenotypes. The second is a subset-based meta-analysis *ASSET* [Bhattacharjee et al., 2012], which provides an optimal subset of non-null traits. And the third is a modified Benjamini-Hochberg (B-H) procedure of controlling the expected false discovery rate [Benjamini and Hochberg, 1995] in the framework of phenome-wide association study. From our simulations we see that an inverse regression based approach *MultiPhen* [O'Reilly et al., 2012] is more powerful than *ASSET* for detecting overall pleiotropic association, except for when all the phenotypes are associated and have genetic effects in the same direction. For determining which specific traits are pleiotropic, the modified B-H procedure performs consistently better than the other two methods. The inverse regression based selection methods perform competitively with the modified B-H procedure only when the phenotypes are weakly correlated. The efficiency of *ASSET* is observed to lie below and in between the efficiency of the other two methods when the traits are weakly and strongly correlated, respectively. In our application to a large GWAS, we find that the modified B-H procedure also performs well, indicating that this may be an optimal approach for determining the traits underlying a pleiotropic signal.

Keywords

Pleiotropy; multiple phenotypes; multivariate association; selection; non-null traits

Introduction

Evaluating the potential pleiotropic impact of genetic variants on multiple traits can improve power to detect association and provide important biological insights [Klei et al., 2008; O'Reilly et al., 2012; Stephens, 2013; Zhu et al., 2014]. A number of methods exist for assessing pleiotropy, and these generally test for overall evidence of pleiotropy across multiple traits [Galesloot et al., 2014]. Once pleiotropy is detected, however, few approaches can distinguish which specific traits underlie the association. Identifying the optimal subset

* *Address for correspondence:* John Witte, Department of Epidemiology and Biostatistics, University of California, San Francisco, 1450 3rd St, Box 3110, San Francisco, CA 94158, JWitte@ucsf.edu.

The authors do not have any conflict of interest.

of traits associated with a particular genetic variant may help uncover shared biological mechanisms across traits. If the phenotypes are seemingly heterogeneous, determining the subset of non-null traits is crucial to better understand the pleiotropy signal.

For example, the levels of low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, and total cholesterol are important blood lipid traits that may impact coronary artery disease. The Global Lipids Genetics Consortium [2013] discovered novel pleiotropic loci associated with different subsets of the traits. For example, variants in the genes *RSPO3, FTO, VEGFA, PEPD* were associated with HDL and triglycerides, but not with total cholesterol or LDL.

Here we explore three approaches that can be used to determine pleiotropic loci that impact subsets of traits: 1) model selection or shrinkage methods; 2) a subset-based meta-analysis approach across phenotypes; and 3) a sequential procedure controlling the false discovery rate. Model selection or shrinkage methods can be applied in the framework of inverse regression of genotype on phenotypes to select an optimal subset of non-null traits corresponding to a SNP associated with multiple traits. One can use *ASSET* [Bhattacharjee et al., 2012] to undertake a subset-based meta-analysis to distinguish an optimal subset of non-null traits. And finally, one can also use a modified Benjamini-Hochberg (B-H) procedure of controlling the expected false discovery rate [Benjamini and Hochberg, 1995] in the framework of phenome-wide association studies (PheWAS) that regresses the individual phenotypes on the SNP genotypes. We evaluate and compare these three approaches via simulation and application.

In the inverse regression framework [O'Reilly et al., 2012; Yan et al., 2013; Wang, 2014; Wu and Pankow, 2015; Majumdar et al., 2015], one can use various model selection techniques, including Akaike information criterion (*AIC*) [Akaike, 1992], Bayesian information criterion (*BIC*) [Schwarz et al., 1978], and extended *BIC* [Chen and Chen, 2008, 2012]. A key advantage of inverse regression is that the distribution of the phenotypes can be arbitrary, allowing the flexibility of simultaneously including both discrete and continuous traits in the analysis. To test for multivariate association, O'Reilly et al. [2012] introduced *MultiPhen* based on proportional odds logistic regression (POLR) of genotype on phenotypes.

With the likelihood framework underlying POLR, one can employ a model selection criterion that explores all possible subsets of traits to find the set which minimizes the expected loss of information penalized by a measure of the model complexity. Under the same set-up, one can also implement the least absolute shrinkage and selection operator (LASSO) [Tibshirani, 1996] (or the adaptive lasso [Zou, 2006]), which shrinks the regression parameters for less important predictors to zero and draws an inference on the sparsity (null traits in our context). The adaptive LASSO has superior performance than the original LASSO [Tibshirani, 1996]. However, there is limited software for implementing LASSO in the POLR framework. One can instead consider the regression of the allelic status on phenotypes [Majumdar et al., 2015] which assumes Hardy-Weinberg equilibrium (HWE) conditioning on the phenotype values and uses the likelihood of a simple logistic regression (LR).

With *ASSET* [Bhattacharjee et al., 2012] one constructs a test statistic to evaluate pleiotropic association based on maximizing a weighted sum of univariate normalized association statistics across all possible subsets of traits in the framework of fixed-effects meta analysis. Even though *ASSET* was primarily designed for multiple studies of distinct case-control phenotypes, it can be implemented for multiple traits in a cohort. A key requirement for using *ASSET* is to provide the correlation matrix of the trait-specific association statistics. The explicit expression of this matrix may be difficult to obtain for multiple phenotypes of different types. Hence, in the methods section we discuss an alternative way of estimating the correlation matrix using the approximate relation between the score function and maximum likelihood estimate (MLE) of a regression parameter.

To implement the modified B-H procedure at a SNP associated with multiple phenotypes, one can first compute the p-value of association for each individual phenotype by regressing it on the SNP genotype, and then applying the modified B-H rule on the trait-specific univariate p-values. The false discovery rate (FDR) and true discovery rate (TDR) are defined as the proportion of null traits and non-null traits among the associated phenotypes, respectively. The advantage of the B-H procedure is that it can keep the expectation of FDR at a desired level. The accuracy of selection of non-null traits is measured in terms of specificity and sensitivity, where specificity is the proportion of null traits discarded from the optimal subset and sensitivity is the proportion of non-null traits included in the subset. Clearly, higher specificity implies lower FDR, and higher sensitivity implies higher TDR. The exact relations between these different notions of true and false selection rates are discussed in the Supplementary materials.

An important property of these three approaches for distinguishing a subset of non-null traits underlying a pleiotropy signal is their ability to incorporate both categorical and continuous phenotypes. Moreover, all of the methods are computationally feasible and suitable to adjusting for potential covariates such as estimates of population structure.

In this paper, we first evaluate via simulation the relative power to detect overall pleiotropic association using *MultiPhen* [O'Reilly et al., 2012], *BAMP* [Majumdar et al., 2015], and *ASSET*. We then compare the accuracy of selection of non-null traits by simulations for the three approaches discussed above: model selection and shrinkage methods based on inverse regression, *ASSET*, and a modified B-H procedure in the PheWAS framework. We consider a range of simulation scenarios for both multiple continuous and binary traits. Finally we apply these methods to the large "Resource for Genetic Epidemiology Research on Adult Health and Aging" (GERA) Kaiser cohort [dbGaP Study Accession: phs000674.v1.p1], evaluating the potential genome-wide (GW) pleiotropy underlying risk of hypertension, Type II diabetes, cardiovascular disease, and cancers.

Methods

Let $Y = (Y_1, \ldots, Y_K)$ denote a multivariate phenotype vector, where each Y_j can be either a categorical or a continuous quantitative trait. And let X denote a vector of SNP genotypes at the marker locus coded as the number of minor alleles (0, 1, 2). For the i^{th} individual, $i = 1$,

…, n, $Y_i = (Y_{i1}, \ldots, Y_{iK})$ and X_i are the multivariate phenotype and minor allele count at the marker locus, respectively.

Inverse regression of genotype on phenotypes

Let $f(Y, X)$ denote the joint probability distribution of Y and X. While regressing phenotype on genotype, $f(Y, X)$ can be decomposed as $f(Y|X) \times f(X)$, where $f(Y|X)$ is the multivariate distribution of the phenotypes conditioning on the genotype. However, modeling $f(Y|X)$ may be very difficult for correlated traits that have non-normal distributions or different forms (e.g., categorical and continuous). A more flexible framework entails inverse regression of genotype on phenotypes whereby $f(Y, X)$ is decomposed as $f(X|Y) \times f(Y)$. Since, $f(X|Y)$ contains the regression parameters that reflect the association between phenotypes and genotype, $f(Y)$ need not be specified in a likelihood ratio test (LRT) for association.

Proportional Odds Regression—One can model $f(X|Y)$ with proportional odds logistic regression (POLR) [O'Reilly et al., 2012] as:

$$log \frac{P(X > x|Y = y)}{P(X \leq x|Y = y)} = \alpha_x + \sum_{j=1}^{K} \beta_j y_j, x = 0, 1.$$

Thus, the conditional distribution of the genotypes given $Y = y$ is:

$$P(X=0|y) = \frac{1}{1+exp\left(\alpha_0 + \sum_{j=1}^{K} \beta_j y_j\right)}, P(X=1|y) = \frac{1}{1+exp\left(\alpha_1 + \sum_{j=1}^{K} \beta_j y_j\right)} - \frac{1}{1+exp\left(\alpha_0 + \sum_{j=1}^{K} \beta_j y_j\right)},$$

$$P(X=2|y) = \frac{exp\left(\alpha_1 + \sum_{j=1}^{K} \beta_j y_j\right)}{1+exp\left(\alpha_1 + \sum_{j=1}^{K} \beta_j y_j\right)},$$

with the restriction that $\alpha_0 \geq \alpha_1$ (in order to ensure $P(X = 1|y)$ is non-negative). Based on the POLR, a LRT can be performed to test $H_0 : \beta_1 = \ldots = \beta_K = 0$ versus H_1 : at least one $\beta_j \neq 0$, $j = 1, \ldots, K$. O'Reilly et al. [2012] termed this method *MultiPhen*.

Allelic Modification—The conditional distribution of X (the number of minor alleles) given $Y = y$ can also be modeled as Binomial with parameters 2 and $p(y)$, where $p(y)$ is the logistic link function [Majumdar et al., 2014, 2015] given by:

$$p(y) = \frac{exp\left(\alpha + \sum_{j=1}^{K} \beta_j y_j\right)}{1+exp\left(\alpha + \sum_{j=1}^{K} \beta_j y_j\right)}.$$

It follows that $p(y) = P(\text{an allele at the marker locus is minor}|y)$. The model implicitly assumes that, conditioned on the vector of multivariate phenotype, the marker locus is in Hardy-Weinberg Equilibrium (HWE). It considers a single intercept parameter α (instead of α_0 and α_1 in *MultiPhen*) and is based on the simple likelihood of the logistic regression. The

test for association is equivalent to testing $H_0 : \beta_1 = \ldots = \beta_K = 0$ versus H_1 : at least one $\beta_j \neq 0$, $j = 1,\ldots, K$. Under H_0, the minor allele frequency (MAF) at the marker locus is independent of y and is given by: $\frac{exp(\alpha)}{1+exp(\alpha)}$. This model is an allelic modification of *MultiPhen* termed *BAMP* (Binomial regression-based Association of Multivariate Phenotypes) [Majumdar et al., 2015].

Model selection criteria

We describe the different model selection criteria that can be implemented based on the likelihood framework underlying *MultiPhen* or *BAMP* in order to select an optimal subset of non-null traits. While defining the criteria, the likelihood in POLR underlying *MultiPhen* is considered. Let S_k be a subset comprising k traits: $Y_{j_1}, \ldots, Y_{j_k}$, where $j_1,\ldots, j_k \in \{1, \ldots, K\}$. Suppose, $l(S_k) = \prod_{i=1}^{n} f(X_i | Y_{ij_1}, \ldots, Y_{ij_k})$. $l(S_k)$ is a function of $(k + 2)$ parameters: $\alpha_0, \alpha_1, \beta_{j_1}, \ldots, \beta_{j_k}$. Let $L(S_k) = max_{\alpha_0,\alpha_1,\beta_{j_1},\ldots,\beta_{j_k}} l(S_k)$. While developing the different model selection criteria, $-2\log L(S_k)$ is considered as the primary loss (incurred due to selecting S_k) function which is penalized by a measure of the model (S_k) complexity. The subset of traits that attains the minimum value of a model selection criterion is selected as the optimal subset of non-null traits.

The *AIC* [Akaike, 1992] is a decision-theoretic criterion of model selection which was derived based on maximum-likelihood theory. In our context, the expression of *AIC* is given by:

$$-2 \times logL(S_k) + 2 \times (k+2).$$

The *BIC* was motivated by a Bayesian approach that assumes a uniform prior over all possible models and has the following form:

$$-2 \times logL(S_k) + (k+2) \times log(n),$$

where n is the sample size. The expressions of *AIC* and *BIC* differ only in the factor $\frac{1}{2}log(n)$ included in the penalty term of the latter. Thus, *BIC* penalizes a subset with a larger number of traits more heavily than *AIC*. As a consequence, *BIC* tends to select fewer traits than *AIC*.

Due to placing a uniform prior across all possible models, the prior probability assigned to the space of models, each comprising k traits, increases proportionally with the size of the space: $\binom{K}{k}$. The difference between the prior probabilities assigned to the different spaces of models with a fixed number of traits becomes pronounced as K increases. To rectify this bias induced by *BIC*, Chen and Chen [2008] considered a prior that assigns equal prior probabilities to each space of models containing a fixed number of traits. This extended Bayesian information criterion (extended *BIC*, also denoted by BIC_γ), is given by:

$$-2 \times logL\,(S_k) + (k+2) \times log\,(n) + 2\gamma \times log \begin{pmatrix} K \\ k \end{pmatrix}, 0 \leq \gamma \leq 1.$$

Clearly, for $\gamma = 0$, $BIC_\gamma = BIC$. In our context, we consider two choices of γ: 0.5 and 1.

Adaptive LASSO

LASSO [Tibshirani, 1996] puts an upper bound on the absolute sum of the main regression parameters which leads to a shrinkage of the parameters corresponding to less important predictors to zero. In our context, LASSO maximizes the penalized log-likelihood: $\sum_{i=1}^{n} logf\,(X_i|Y_i) - \lambda \sum_{j=1}^{K} |\beta_j|$ with respect to $\beta_1,\ldots,\beta_K$, where λ is the tuning parameter controlling the amount of shrinkage. Hence, $\lambda = 0$ induces no shrinkage, and selects all traits due to maximizing the unrestricted likelihood. On the other hand, a sufficiently large value of λ will lead to exclusion of all traits.

To induce differential shrinkage across βs, adaptive LASSO (A-LASSO) [Zou, 2006] considers a weight $\frac{1}{|\beta_{j,\text{free}}|}$ and maximizes the penalized log-likelihood: $\sum_{i=1}^{n} logf\,(X_i|Y_i) - \lambda \sum_{j=1}^{K} \frac{|\beta_j|}{|\beta_{j,\text{free}}|}$, where $\beta_{j,\text{free}}$ is the maximum likelihood estimate (m.l.e.) of β_j in the unrestricted likelihood. A-LASSO is widely used because of its superior performance over the basic LASSO with respect to selection efficiency.

While implementing A-LASSO, we use the likelihood framework underlying *BAMP* with the R package *glmnet* [Friedman et al., 2009] to track the entire regularization path. A key step is to select an optimal λ among its values along the regularization path. We employ both the *AIC* and *BIC* in which the number of non-zero βs is treated as the degree of freedom [Zou et al., 2007] to select an optimal λ. For a selected λ, the subset of non-null traits is constructed by excluding Y_j from (Y_1, …, Y_K) if β_j is zero, $j = 1, \ldots, K$.

ASSET—For analyzing multiple phenotypes, Bhattacharjee et al. [2012] introduced a novel method *ASSET* which simultaneously provides the p-value of pleiotropic association and an optimal subset of non-null traits. We briefly describe the theoretical framework underlying *ASSET* which is based on the set-up of regressing phenotypes on genotype. While regressing Y_k on X by a linear or logistic regression as appropriate, let $\hat{\beta}_k, v_k$ be the estimates of the association parameter and its standard error, $k = 1, \ldots, K$.

Adopting the framework of a fixed-effects meta analysis, for a subset of traits S, *ASSET* defines the Z statistic as: $Z\,(S) = \sum_{Y_k \in S} \sqrt{\pi_k\,(S)} Z_k$, where $Z_k = \frac{\hat{\beta}_k}{v_k}$, and $\pi_k(S)$ is an appropriate weight associated with Y_k belonging to the subset S. For example, there are K separate GWAS for Y_1, …, Y_K with the k^{th} study having a sample size n_k. Then, $\pi_k\,(S) = \frac{n_k}{\sum_{k \in S} n_k}$. The global association of a SNP with at least one of the traits is

measured by the test-statistic: $Z_{max}=max_{S\in\mathscr{A}}|Z(S)|$, where $\mathscr{A}$ is the space of all possible S. In addition to the p-value of global association, *ASSET* also offers an optimal subset of non-null traits that are associated with the SNP, which is essentially the subset of traits that constructs Z_{max}.

Even though *ASSET* was originally developed for separate case-control studies, it can also be used for multiple phenotypes in a cohort study. A key requirement to implement *ASSET* is the correlation matrix of $\hat{\beta}_k$s under the null hypothesis of no global association. Bhattacharjee et al. [2012] provide an explicit expression of this matrix for multiple case-control phenotypes, which is also straightforward to derive for normally distributed traits. However, for non-normal correlated traits, including different types of traits (e.g., one discrete and one continuous), it may be very difficult to directly obtain the correlation matrix. Since, in the majority of the cases, $\hat{\beta}$ is the MLE of β, we use the approximate relation between the score function and MLE of β in order to estimate the correlation matrix. The score function can be computed for a well-defined probability distribution of a phenotype conditioned on the genotype, which indicates the general applicability of this strategy (described in the Appendix).

Modified B-H procedure

Benjamini and Hochberg [1995] introduced a sequential procedure that controls the expected FDR for multiple hypothesis testing. Suppose, $Y_1, \ldots, Y_K$ are individually regressed on X by linear/logistic regression as appropriate, and let $P_1, \ldots, P_K$ denote the trait-specific p-values of association. Suppose, $P_{(1)}, \ldots, P_{(K)}$ denote the p-values arranged in increasing order of magnitude and let $Y_{j_1}, \ldots, Y_{j_K}$ be the actual phenotypes corresponding to the ordered p-values. Given that w is the desired level of the expected FDR, the B-H procedure states that: if k is the largest l such that $P_{(l)} \leq \frac{l}{K}w$, then declare $Y_{j_1}, \ldots, Y_{j_K}$ to be associated (non-null).

Since the procedure may conclude that none of the phenotypes is associated even though an overall test detected pleiotropy, to ensure that the optimal subset of non-null traits contains at least one phenotype, we consider a modified level of the expected FDR as: $\alpha_w^*=max\left(w, KP_{(1)}\right)$, $0 < w < 1$. Now, the optimal subset will always contain the trait with the minimum trait-specific univariate p-value.

Benjamini and Hochberg [1995] proved that the sequential procedure controls the expected FDR when the test statistics are independent. Later, Benjamini and Yekutieli [2001] proved that, for dependent test statistics, which is the case for correlated traits, the same procedure controls the expected FDR when the test statistics have positive regression dependency on each of the test statistics corresponding to the true null hypotheses [Benjamini and Yekutieli, 2001]. They also showed that this property is satisfied in the scenario when the test statistics are distributed as multivariate normal with a positive definite correlation matrix. In our context, the test statistics in the regression of individual phenotype on genotype is normally distributed. Hence, under the assumption that the joint distribution of the individual test

statistics is multivariate normal with a positive definite correlation matrix, the B-H procedure should control the expected FDR at a desired level.

Simulation study

First we carry out simulations to compare *MultiPhen*, *BAMP*, and *ASSET* with respect to the power of detecting overall pleiotropic association. Next, we evaluate the efficiency of the above three strategies to determine the subset of non-null traits underlying a signal of multivariate association. The accuracy of the selection of non-null traits is measured by specificity and sensitivity.

Simulation models

We assume that a bi-allelic quantitative trait locus (QTL) with alleles A (minor) and a has genetic effect on Y. A bi-allelic marker locus with alleles B (minor) and b is considered such that the coefficient of linkage disequilibrium between the QTL and the marker locus is δ, which is defined as: $\delta = P(AB) - P(A)P(B)$. Let the frequencies of the minor alleles A and B be denoted by p and m, respectively.

We describe the models to simulate the data on genotypes at the QTL and the marker locus, and the vector of multiple phenotypes. HWE is assumed at both of the loci while simulating the genotype data. Let G denote the genotype at the QTL. Thus, $G = AA, Aa, aa$, and under HWE, $P(AA) = p^2$, $P(Aa) = 2p(1-p)$, $P(aa) = (1-p)^2$. Using the MAF at the marker locus (m) and its LD with the QTL (δ), the genotype at the marker locus is generated conditioning on the genotype at the QTL.

We use an additive model to simulate the vector of multiple phenotypes following Galesloot et al. [2014]. Let g denote the count of minor alleles at the QTL. For a given $j \in \{1, \ldots, K\}$, the following model is assumed: $Y_j = \beta_j \times g + e_j$, where $g = 0, 1, 2$. Irrespective of the QTL genotype, the random residual e_j is distributed as $N\left(0, \sigma^2_{ej}\right)$. Thus, $E(Y_j|AA) = 2\beta_j$, $E(Y_j|Aa) = \beta_j$, and $E(Y_j|aa) = 0$.

Let the total variance of Y_j be denoted by σ^2_j. Of note, $V(Y_j) = V(E(Y_j|G)) + E(V(Y_j|G))$, and under the above model, $V(E(Y_j|G)) = \beta^2_j \times 2p(1-p)$. Since, $V(Y_j|G) = \sigma^2_{ej}$, irrespective of the QTL genotype G, $E(V(Y_j|G)) = \sigma^2_{ej}$. Hence, $\sigma^2_j = \beta^2_j \times 2p(1-p) + \sigma^2_{ej}$.

Let $h^2_j\ (\in (0,1))$ denote the trait-specific heritability of Y_j due to the QTL. Hence, $h^2_j = \frac{V(E(Y|G))}{\sigma^2_j}$. Thus, for a given choice of h^2_j, the residual variance is given by: $\sigma^2_{ej} = \left(1 - h^2_j\right)\sigma^2_j$. Since, $V(E(Y|G)) = h^2_j \times \sigma^2_j$, for a given choice of $\left(h^2_j, \sigma^2_j, p\right)$, $\beta_j = \sqrt{\frac{h^2_j \sigma^2_j}{2p(1-p)}}$.

From the above model, the genetic correlation between a pair of traits Y_j and Y_k is positive or negative depending on the signs of β_j and β_k. We consider that $e_1, \ldots, e_K$ follow a

multivariate normal distribution with a correlation matrix $((\rho_{e,jk}))$. For $j \neq k$, ρ_{ejk} is the correlation between e_j and e_k, the random residuals of Y_j and Y_k excluding the QTL effect. Thus, the set of parameters that completely specifies the simulation model for generating data on the genotypes and multiple phenotypes is given by: $\left(p, m, \delta, h_j^2, \sigma_j^2, \rho_{e,jk}\right)$, for $j, k \in \{1,\ldots, K\}$.

To investigate non-normal continuous traits, the random residual e_j is assumed to be distributed as chi-square with 1 d.f.. We assume $e_j = E_j^2$, where $E_1, \ldots, E_K$ follow multivariate normal with covariance matrix $((\rho_{E,jk}))$, so $V(E_j) = 1$ and $Cov(E_j, E_k) = \rho_{E,jk}$. Under this set-up, for a given choice of h_j^2 and p, $\beta_j = \sqrt{\frac{h_j^2}{\left(1-h_j^2\right) p\,(1-p)}}$. For simplicity's sake, we consider the same choice of $((\rho_{e,jk}))$ and $((\rho_{E,jk}))$ in our simulation study.

Design

Power comparison—*MultiPhen*, *BAMP*, and *ASSET* are compared with respect to the power of detecting multivariate association with a vector of four (K) phenotypes. We implement *ASSET* with the option of a one-sided search when all $\hat{\beta}_S$ are positive and a two-sided search when $\hat{\beta}_S$ are both positive and negative. The rate of type-I error and power are estimated at the level of significance 0.01 based on a sample of $10K$ individuals and $5K$ replications.

Various simulation scenarios are explored to study the difference in the power of the three methods. The trait-specific heritabilities due to the QTL are considered to have the same – and then different – magnitudes, including the special case of oppositely directed QTL effects across traits. All choices of the trait-specific heritabilities are listed in Table 1. We denote the residual correlation structure among the phenotypes by $r_E = \left(r_{E_1}, r_{E_2}\right)$, where r_{E_1} denotes the residual correlation between a pair of non-null traits, and r_{E_2} denotes the residual correlation between a pair of null traits or one null and one non-null trait. The following different choices of r_E are considered: (0.05, 0.05), (0.3, 0.3), (0.5, 0.5).

Without loss of generality, the first K_1 and last K_0 of the total K traits are assumed to be associated and non-associated with the pleiotropic locus, respectively. To consider binary traits, every trait of the vector of continuous traits is dichotomized above a threshold subject to fixing the prevalence of the binary trait at 10%.

A MAF = 0.1 is assumed at both of the marker locus and QTL. Different levels of the standardized LD between the QTL and marker locus are considered. LD = zero corresponds to the type I error rate for the methods. We investigate the situations in which all – and then a subset of the phenotypes are non-null, respectively.

Selection of non-null traits—The three different strategies (described in the "Methods" section) are compared with respect to the accuracy of selecting non-null traits underlying a pleiotropic association. We consider various simulation scenarios of

phenotypes and trait-specific heritability which are described in Table 3. The following different choices of r_E are considered: (0.05, 0.05), (0.3, 0.1), (0.3, 0.3), (0.5, 0.3), (0.5, 0.5). A high value of the standardized LD (0.95) between the QTL and marker locus is assumed. The specificity and sensitivity are estimated based on 500 replications, except for Tables S5, S6, S7, S8, where 200 replications were used to reduce computing time.

Since our goal is to determine the most efficient and accurate approach to identifying the non-null traits underlying a genome-wide signal of pleiotropic association, we ascertain in every replication whether the multivariate association p-value $< 5 \times 10^{-8}$. For continuous traits, a sample size of $15K$ individuals is used while carrying out the simulations. Since the dichotomization of continuous traits reduces power, for binary traits a larger sample size of $30K$ individuals is used. A MAF = 0.1 is assumed at the QTL and MAFs of 0.1 and 0.2 at the marker locus are considered.

We primarily consider four or eight phenotypes. We also consider 20 phenotypes in Table S8. With 20 phenotypes, more than $1M$ possible subsets of traits have to be explored for *AIC*, *BIC*, and extended *BIC*. To reduce this number, we minimized these criteria only for the sequence of subsets appearing along the entire regularization path of adaptive LASSO. This strategy was also adopted in Chen and Chen [2008, 2012]. Note that, *LSA*, *LSB*, *ASSET*, or MBH_w are computationally feasible for a large number of phenotypes.

Results

Power comparison—All three methods maintain appropriate false positive rates (Table 2), and as expected the power increases as the LD increases between the marker locus and the QTL. In general, the power to detect pleiotropy is higher when the QTL effects on the associated traits are both positive and negative than when the QTL effects are all positive (Figure 1, S1, S2).

Power plots for continuous normal traits are presented in Figure 1, and the power plots for binary traits are presented in Figure S1. The power plots for continuous traits distributed as chi-square are given in Figure S2. The power curves show that *BAMP* always produces slightly higher power (1%–3%) than *MultiPhen*. When two of the four traits are associated, *MultiPhen* produces 1%–35% higher power than *ASSET* for continuous normal traits (Scenario 1 in Figure 1) and 1%–15% higher power for binary traits (Scenario 3 in Figure S1). When all the four phenotypes are associated and the QTL effects are positive, *ASSET* yields 1%–18% higher power for continuous traits (Scenario 2 (choice 1) and Scenario2 (choice 2) in Figure 1) and 1%–11% higher power for binary traits (Scenario 4 (choice 1) and Scenario 4 (choice 2) in Figure S1); but if the QTL effects are both positive and negative, *MultiPhen* offers 1%–52% higher power for continuous traits (Scenario 2 (choice 3) in Figure 1) and 1%–22% higher power for binary traits (Scenario 4 (choice 3) in Figure S1). For continuous traits distributed as chi-square, the relative performance between *MultiPhen* and *ASSET* remain similar as for normal traits (Figure S2). Of note, the difference in power between the two methods increases with the strength of correlation among phenotypes.

Selection of traits—For the selection criteria, we desire a balance between a good level of both sensitivity (correctly selecting the non-null traits) and specificity (correctly excluding null traits). First we discuss the relative performance of *BIC*, $BIC_{0.5}$, BIC_1, since these are variations of BIC_γ for $\gamma = 0, 0.5, 1$, respectively. In most scenarios, *BIC* performs better than $BIC_{0.5}$ and BIC_1 (Tables S1–S12). For example, *BIC* yields 1%–6% higher specificity than $BIC_{0.5}$ and 1%–14% higher specificity than BIC_1. *BIC* also offers a sensitivity increase of 1%–5% and 1%–10% compared to $BIC_{0.5}$ and BIC_1, respectively. However, in a few cases for eight phenotypes, $BIC_{0.5}$ and BIC_1 perform marginally better than *BIC*. But overall, *BIC* appears to be a better alternative than $BIC_{0.5}$ and BIC_1, and so we focus on *BIC* here.

The modified B-H procedure with the level of expected FDR α_w^* is denoted by MBH_w. In the simulation study, we consider three different choices of w: 0.01, 0.05, and 0.1. $MBH_{0.01}$ produces higher specificity but marginally lower sensitivity than $MBH_{0.05}$ and $MBH_{0.1}$ with the latter offering the highest sensitivity as expected (Tables S1 – S12).

The performance of all selection criteria improves substantially when the QTL effects on the non-null traits are both positive and negative compared to when the QTL effects are all positive (Tables S1 – S12). Hence the plots corresponding to oppositely directed QTL effects are excluded from Figure 2 which corresponds to numerical results given in Tables S2 – S4.

The selection criteria based on the inverse regression perform better for weakly correlated traits ($r_E = (0.05, 0.05)$) than for traits having stronger correlation structure (Figure 2). The specificity decreases as the strength of correlation increases, especially for, *AIC* and the A-LASSO using *AIC* to select an optimal shrinkage parameter (*LSA*). As expected, the specificity is better for $r_E = (0.3, 0.1)$ when the null and non-null traits are differentially correlated than for $r_E = (0.3, 0.3)$ when the null and non-null traits are equally correlated. A similar pattern is observed between $r_E = (0.5, 0.3)$ and $r_E = (0.5, 0.5)$. *AIC* always produces higher sensitivity than *BIC* at the cost of less specificity (Figure 2). Similarly, *LSA* offers higher sensitivity than the A-LASSO using *BIC* to select an optimal shrinkage parameter (*LSB*) at the expense of less specificity (Figures 2). For example, in Scenario 7 of Figure 2, *AIC* yields 1%–35% higher sensitivity than *BIC* at the cost of 1%–56% lower specificity, and *LSA* offers 1%–20% higher sensitivity than *LSB* at the expense of 1%–42% lower specificity.

ASSET offers a moderate level of specificity that remains fairly stable across different degrees of correlations among traits. However, its sensitivity decreases as the strength of correlation among phenotypes increases (Figure 2). For weakly correlated traits ($r_E = (0.05, 0.05)$), the inverse regression approach performs better than *ASSET*. But, *ASSET* maintains better specificity than the inverse regression based criteria for phenotypes with stronger correlation.

The modified B-H procedure ($MBH_{0.01}$) outperforms the other methods with respect to good levels of both the specificity and sensitivity, in particular, when the phenotypes have higher correlation (Figure 2). The inverse regression approach performs competitively only when the phenotypes are weakly correlated ($r_E = (0.05, 0.05)$).

We also performed simulations encompassing various other scenarios. For example, in Table S5, eight continuous traits are distributed as chi-square among which four are non-null. In Tables S6 and S7, three and six out of eight normally distributed traits are associated, respectively. In these simulations, the relative behavior of the methods considered here were similar to that described above. Table S8 gives results for twenty normal traits, of which eight are non-null. Here, *ASSET* performs fairly well across different scenarios, but is consistently outperformed by MBH_w. In contrast, the selection criteria based on the inverse regression of genotype on phenotypes perform very poorly as the phenotypic correlation increases (Table S8).

Application to Cohort Study

To demonstrate the performance of the different methods based on real data, we analyzed the large genome-wide association study "Resource for Genetic Epidemiology Research on Adult Health and Aging" (GERA) cohort data obtained from dbGaP [dbGaP Study Accession: phs000674.v1.p1]. For simplicity's sake, we restricted our comparison to 62,318 European-American individuals who constitute more than 75% of the dbGaP data. The data include 657,184 SNPs genotyped across 22 autosomal chromosomes. We consider four binary phenotypes: hypertension (HYP), Type II diabetes (T2D), cardiovascular disease (CVD), and any cancer (CAN). The phenotypes in the GERA cohort were obtained using electronic health record (EHR). Note that in the dbGaP GERA cohort data, the cancers were collapsed into a single variable (any cancer). Therefore, we could only use an overall cancer categorization even though the genetic architecture is likely heterogeneous across different cancers.

Before our analysis, we undertook the following QC steps. First, we removed individuals with: over 3% of genotypes missing; any missing information on covariates (described below); genotype heterozygosity outside six standard deviations; first degree relatives; or discordant sex information. This leaves us with 53,809 individuals. Next, we removed SNPs with: MAF < 0.01; 10% or more missingness; or deviation from HWE at a level of significance 10^{-5}. This leaves 601, 335 SNPs which were tested for pleiotropic association by *MultiPhen* and *ASSET*. However, *MultiPhen* failed to converge for 4, 282 SNPs. So different methods were finally applied to 597,053 SNPs. We adjust all of the analyses for the following covariates: age, gender, smoking status, BMI category, and six principal components of ancestry (PCs).

The overall correlation matrix of the four phenotypes estimated based on the individuals analyzed here is given in Table 4. As expected, the highest correlation was between hypertension and CVD. Also, the pairwise correlations are non-negligible except between T2D and any cancer.

The single phenotype association with each of the four phenotypes is tested by a logistic regression of the case-control status on genotype incorporating the same set of covariates, and the estimated association parameters (βs) are then used for implementing *ASSET* and modified B-H procedure. For every genome wide significant SNP detected by either of the methods, all the three strategies for selecting the optimal subset of non-null traits are

implemented. The Q-Q plots for the p-values of pleiotropic association obtained by the two methods are presented in Figure S4. For *MultiPhen*, the genomic control inflation factor (λ) is estimated as 1.05. However, it is very difficult to estimate λ for *ASSET*. But the Q-Q plot for *ASSET* as well as for *MultiPhen* suggest that the results are appropriately adjusted for covariates.

Results

The MANHATTAN plots for the Kaiser cohort application results from *ASSET* and *MultiPhen* are presented in Figure S3. Since the sets of genome-wide associated SNPs found by *MultiPhen* and *ASSET* overlap substantially, results for the union of the two sets are provided in Tables 5, 7, S13, and S14. We apply the conventional genome wide level of statististical significance 5×10^{-8}. We also report the SNPs that fall just outside this cut-off (5×10^{-8}–6×10^{-8}) for either of the methods. *MultiPhen* identifies 50 significant SNPs and *ASSET* detects 56. Each of the sets of SNPs are filtered for pair-wise LD ($r^2 \geq 0.2$) which leaves us with 14 independent SNPs found by *MultiPhen* and 17 SNPs identified by *ASSET* (Table 5). Two of the SNPs uniquely identified by *ASSET* are actually detected by *MultiPhen* with p-values just below the significance cut-off, chr3:*rs*11709077 (p-value: 5.58×10^{-8}) and chr8:*rs*13266634 (p-value: 5.59×10^{-8}).

MultiPhen also does not identify the following three SNPs detected by *ASSET*: chr5:*rs*183671, chr6:*rs*872071, chr11:*rs*67279079. And *ASSET* misses the SNPs detected by *MultiPhen*: chr10:*rs*7896811 and chr16:*rs*4408545. *ASSET* does, however, detect the other three SNPs detected by *MultiPhen* on chromosome 10 (listed in Table 5). Nevertheless, when either of the methods fails to detect a SNP identified by other method, it still exhibits suggestive evidence of association (a p-value of 10^{-7} order).

We also map the identified SNPs into genes using the human assembly GRCh37/hg19 in the UCSC genome browser. A SNP is assigned to a gene if it is located either within the boundaries of the gene or 20 KB upstream or downstream. The name of the mapped genes and their known associations with various phenotypes are provided in Table 5. We observe from the gene mapping that a substantial proportion of the significant SNPs (13 out of 19) map to at least one gene and two SNPs map to multiple genes (Table 5). If a SNP maps to a gene but falls outside its boundaries, the SNP's distance from the nearest of the two boundaries of the gene is also provided. Out of 13 SNPs mapped to genes, four SNPs did not fall within the boundaries of a gene. Among 9 SNPs which are located within a gene, all four SNPs on chromosome 10 mapped to the same gene *TCF7L2*. Thus these 9 SNPs mapped to six distinct genes. These genes are associated with a range of phenotypes including T2D, CVD, various cancers, and glycated hemoglobin levels which is associated with increased risk of hypertension [Bower et al., 2012]. Of note, *rs*13266634 is within the 8q24 chromosomal region, associated with prostate and other cancers [Witte, 2007].

The selection results of traits underlying the significant SNPs are graphically presented in Tables 7 and S13. Table 7 contains the significant SNPs at which multiple non-null traits were detected by at least two criteria, and Table S13 presents the rest of the results. The univariate association p-values for individual phenotypes are also presented in Table S14. With respect to selection of traits, *BIC* and *LSB* behave conservatively (Tables 7, S13, and

S14). *BIC* selects hypertension and cancers at chr6:*rs*12203592, but selects one phenotype (that has the minimum trait-specific univariate association p-value) at all of the other SNPs. For all the SNPs listed in Table S14, *LSB* selects one trait (that has the minimum univariate association p-value). However, *LSB* selects CVD and cancers at chr9:*rs*10116277 that is listed in Table S14. Of note, *BIC* also selects CVD and cancers at chr9:*rs*9632885, chr9:*rs*1333040 which are in strong LD with chr9:*rs*10116277. *AIC* and *LSA* perform similarly, and select multiple phenotypes as non-null at 8 of the 19 SNPs. *LSA* selects an additional phenotype compared to *AIC* at two SNPs: chr8:*rs*13266634 and chr9:*rs*10116277.

ASSET selects multiple non-null phenotypes at 12 SNPs (Tables 7, S13, and S14). However, the subset of non-null traits selected by *AIC* and *LSA* are substantially more consistent with the trait-specific univariate association p-values (in terms of phenotypes with smaller trait-specific p-values being selected as non-null) compared to *ASSET* (Tables 7, S13, and S14). Of note, in the simulation study for four binary traits, *AIC* and *LSA* performed better than *ASSET* except for the case $r_E = (0.5, 0.5)$ (Figure 2, Tables S2, S10). The correlation matrix of the four case-control phenotypes in Table 4 shows that the correlation between different pairs of diseases varies in the range 0.05–0.33.

Finally $MBH_{0.01}$ and $MBH_{0.05}$ select multiple non-null phenotypes at 4 and 5 SNPs, respectively. As expected, $MBH_{0.05}$ identifies an additional non-null trait compared to $MBH_{0.01}$ at chr3:*rs*1470580, chr6:*rs*12203592, and chr9:*rs*10116277, respectively. $MBH_{0.1}$ identified the same traits at the same 5 SNPs as $MBH_{0.05}$; at chr10:*rs*12255678 it selected cancers in addition to hypertension and T2D identified by $MBH_{0.05}$.

There may be some scenarios in which *AIC*, *LSA*, *ASSET* can select multiple non-null traits, but MBH_w will pick up a single phenotype due to a fixed choice of w. For example, for chr6:*rs*872071, *AIC*, *LSA*, *ASSET* selected hypertension and cancers as non-null traits, but both $MBH_{0.05}$ and $MBH_{0.01}$ picked up only cancers, because the univariate association p-value for cancers is the smallest and the p-values for other traits did not pass the modified B-H procedure (Table 7). The same scenario is also observed for chr10:*rs*7079711, where *AIC*, *LSA*, *ASSET* picked up T2D and cancers, but MBH_w selected only T2D as non-null. Whenever a selection criterion included a single trait, that trait had the smallest trait-specific univariate association p-value.

We compared these pleiotropy results to the previously published GWAS findings that are in LD with our significant SNPs ($r^2 \geq 0.2$) from the NHGRI-EBI catalog (Tables S15 and S16). Since $MBH_{0.01}$ and $MBH_{0.05}$ offered the most reliable selection performance in our simulation study, the phenotypes identified by these criteria are listed in Tables S15 and S16. Sixteen out of 19 SNPs are in LD with at least one known GWAS hit. Among these, every SNP at which single phenotype is selected by both $MBH_{0.01}$ and $MBH_{0.05}$ is in LD with at least one NHGRI-EBI GWAS hit which is associated with either the selected GERA phenotype or a closely related phenotype. For example, only cancers was selected at chr2:*rs*3769823 which is in LD with two NHGRI-EBI GWAS hits that are known to be associated with chronic lymphocytic leukemia, melanoma, and esophageal squamous cell carcinoma. Pleiotropic loci also exhibited such consistency with previous results. For example, HYP and T2D were selected by both of the criteria at chr10:*rs*12255678 and

chr10:*rs*4506565, which are in LD with NHGRI-EBI GWAS hits that are associated with T2D and glycated hemoglobin levels – a risk factor for hypertension [Bower et al., 2012]. Overall, while there is substantial overlap between the SNPs detected here and known associations from NHGRI-EBI GWAS catalog, $MBH_{0.01}$ showed more overlap compared to $MBH_{0.05}$ (Tables S15 and S16). Moreover, SNPs not exhibiting such overlap were either potential eQTLs in disease relevant tissues or in known risk genes (chr10:*rs*7079711 is in *TCF7L2*, a T2D gene).

Discussion

Most of the methods for assessing pleiotropic association between a genetic variant and a vector of multiple phenotypes evaluate the overall evidence of association, but do not explore the subset of underlying non-null phenotypes. Here, we have investigated three different strategies to address this objective.

ASSET simultaneously provides a non-null subset of traits along with the global p-value of pleiotropic association, but was previously evaluated for multiple case-control association studies [Bhattacharjee et al., 2012]. For a cohort study of multiple continuous or binary phenotypes, we have compared it with *MultiPhen* with respect to the power of detecting multivariate association. In our simulation study, *MultiPhen* was more powerful than *ASSET* except for the scenario when all the phenotypes are associated and the QTL effects are in the same direction. However, in the application to Kaiser data, *ASSET* detected a slightly larger number of significant signals. *BAMP* consistently produced slightly higher power than *MultiPhen*, as shown in more details in Majumdar et al. [2015].

The finding that *MultiPhen* offers higher power than *ASSET* when two of the four traits are null is counter-intuitive, because *ASSET* was designed to yield higher power in such situations as its test statistic is obtained by maximization across different possible subsets of traits to separate out the non-null traits. One possible explanation for this is that *MultiPhen* models the individual-level data for multiple phenotypes jointly, and hence is more multivariate in nature than *ASSET*, which combines the univariate association statistics. Furthermore, the power and type I error of *ASSET* decrease as the phenotypic correlation increases (r_E = (0.3, 0.3), (0.5, 0.5)) (Figures 1, S1, S2, and Table 2). This conservative behavior of *ASSET* may be due to estimating the p-value of overall pleiotropic association by using the discrete local maxima (DLM) method [Taylor et al., 2007] to approximate tail probabilities of a test statistic that is maximized over a grid.

For the selection of non-null traits, our simulation study shows that the modified B-H procedure performs the best across a range of scenarios considered here. The inverse regression based criteria perform competitively with the modified B-H procedure only when the phenotypes are weakly correlated. The efficiency of *ASSET* is observed to lie below and in between the efficiency of these other two methods when the traits are weakly and strongly correlated, respectively. The issue of multi-colinearity may explain the poor performance of the inverse regression based criteria for strongly correlated phenotypes. The modified B-H procedure offers overall good specificity as it is designed to keep control over the expected FDR and lower FDR induces higher specificity.

In our application, $MBH_{0.01}$ and $MBH_{0.05}$ detected multiple non-null traits at fewer SNPs compared to *AIC*, *LSA*, or *ASSET*, but this may reflect higher specificity of $MBH_{0.01}$ and $MBH_{0.05}$ as demonstrated in the simulation study. $MBH_{0.1}$ selected cancers at chr10:*rs*12255678 in addition to hypertension and T2D; these latter two phenotypes were also identified by both of $MBH_{0.05}$ and $MBH_{0.01}$. *AIC* and *LSA* also selected these three phenotypes as non-null. At this SNP, however, the univariate association p-value for cancers is 0.058. Thus, since $MBH_{0.1}$ may be liberal with respect to the false positive rate, this pleiotropy signal of cancers is only suggestive. Nevertheless, it is worth noting that Michailidou et al. [2013] reported chr10:*rs*113014168 which is in LD with chr10:*rs*12255678 ($r^2 = 0.21$ based on European population) as significantly associated with breast cancer.

We also carried out simulations to compare the methods with respect to power when the marker locus MAF = 0.2. The relative performance of the methods remain similar to when the MAF = 0.1. While studying the selection efficiency for different methods, we also considered a standardized LD of 0.8 between the QTL and marker locus and found the relative performance of the methods to be similar to when the LD is 0.95.

A-LASSO was implemented based on the likelihood underlying *BAMP* that assumes HWE conditional on the phenotype vector. A simulation study evaluating the possible effects of this assumption on the accuracy of the selection of the non-null traits is included in "Supplementary materials" in the section titled "Effects of deviation from HWE on the selection accuracy". The results show that the effect is negligible for weakly correlated traits. Otherwise, the assumption may lead to a 1%–5% decrease in the specificity while increasing the sensitivity by 1%–4%. In the real application, 5 out of the 19 top hits deviate from HWE at a level of significance 0.05; but when increasing the level of significance to 0.001 none of the 19 hits deviated from HWE (Table 5).

The relative performance of different methods also remain similar regardless of the type and distribution of phenotypes, (e.g., continuous traits having normal and chi-square distribution, and binary phenotypes as considered in the simulation study). However, it is difficult to know in general how a particular method will perform for a non-normal phenotype compared to a normally distributed phenotype. For example, Guo et al. [2015] showed that *MultiPhen* loses power when phenotypes have non-normal distribution compared to normally distributed traits.

While implementing *ASSET* for continuous phenotypes, the distribution of the traits needs to be ascertained to calculate the correlation between $\hat{\beta}_S$ using the strategy outlined in the appendix. One can also apply suitable transformations on the phenotypes to induce normality. It is straightforward to employ inverse regression for phenotypes with arbitrary distributions since they are treated as regressors. *ASSET* allows for different directions of the genetic effects across non-null traits but assumes the genetic effects in a given direction to be the same, whereas, the inverse regression allows for heterogeneous genetic effects. However, implementing the inverse regression approach requires individual-level genotype and phenotype data. In contrast, *ASSET* and the modified B-H procedure can be employed

using both individual-level and summary statistics data. Hence like *ASSET*, the modified B-H procedure can also be applied to separate association studies.

While *ASSET* is a one-step procedure that offers a p-value of association and non-null subset of traits simultaneously, the other two approaches are applied to the significant SNPs in the second step. Of note, all three strategies – inverse regression based selection/shrinkage methods, *ASSET*, and modified B-H procedure can be applied to significant SNPs found by any other method of testing multivariate association.

Note that, we have used the same data to detect the most significant pleiotropic signals and then to identify the optimal non-null traits for a selected SNP. Ideally, to implement MBH_w, while computing the p-value of association between a phenotype and a selected SNP, one should condition on the event: {multivariate association p-value for the selected SNP $< 5 \times 10^{-8}$}. Due to this conditioning, under the null hypothesis of no association, the conditional test statistics would no longer follow a multivariate normal distribution, rather a truncated multivariate normal distribution with the truncated region: {multivariate association p-value $< 5 \times 10^{-8}$}. Permutation based algorithms have to be employed to compute such a conditional p-value as its closed form is analytically intractable. However, to compute such a conditional p-value even in the order of 10^{-4}, an average of 2×10^{11} permutations will be required, which can become computationally infeasible for a large data (as in our application). For a selected SNP, without conditioning, the $\hat{\beta}_S$ across correlated traits should follow multivariate normal, because the fact that this SNP has passed through the genome wide level of significance for pleiotropic association is not incorporated while assuming the joint distribution of $\hat{\beta}_S$. In our simulation studies, we applied the MBH rule on the unconditioned univariate p-values of association only in those replications for which the multivariate association p-value was $< 5 \times 10^{-8}$. Along with the specificity and sensitivity, we also estimated the FDR and TDR (results not provided for brevity) and observed that FDR was controlled at the desired levels (0.01, 0.05, or 0.1). This indeed supports our viewpoint that, without conditioning, $\hat{\beta}_S$ for a selected SNP do not deviate from multivariate normality leading to the control of the FDR at a desired level under dependence structure.

The evidence that a pleiotropic signal obtained by MBH_w is true would be stronger as w in MBH_w becomes smaller; however, this is at the expense of discovering smaller number of signals associated with multiple non-null traits (higher specificity at the expense of lower sensitivity). In our simulation study, compared to other criteria, $MBH_{0.01}$ provided consistently better level of specificity and good level of sensitivity across various scenarios that supports its usefulness for selection. $MBH_{0.01}$ also showed substantial overlap between the genome wide signal of associations detected in the GERA cohort and known associations from NHGRI-EBI GWAS catalog. Instead of implementing the MBH rule based on the univariate p-values, we also applied it on two other classes of p-values. First, we calculate the p-value for individual phenotypes based on the $\hat{\beta}_S$ jointly estimated while maximizing the likelihood underlying *MultiPhen*. These $\hat{\beta}_S$ are obtained from the regression of genotype on all phenotypes simultaneously, and hence a $\hat{\beta}$ corresponding to a trait should be adjusted for the effects of the other traits. Second, each phenotype is regressed on the genotype while adjusting for the remaining set of phenotypes as covariates, and a p-value of

genetic association is computed. However, when we applied the MBH rule on these two classes of p-values we found that they perform poorly, producing substantially lower specificity compared to the MBH rule implemented on the univariate p-values of association.

MultiPhen is based on the proportional odds logistic regression, and as implemented in R sometimes failed to converge. When this occurred, the non-convergence warning said that the fitted probabilities were very close to 0 or 1. Nevertheless, *MultiPhen* failed to coverage for a very small proportion of SNPs (0.7%), so this issue did not compromise our objective of evaluating different methods.

Note that another model-based approach for pleiotropy analysis proposed by Stephens [2013] is based on the posterior evaluation of partitions of phenotypes distributing into three categories: direct, indirect, or no association. But, one major limitation of this method is that its implementation is mainly restricted to normally distributed phenotypes. Peterson et al. [2016] proposed a hierarchical B-H procedure that can keep control over the FDR first across the identified SNPs associated with multiple phenotypes and then across the traits within the family of selected SNPs. The procedure is mainly based on the univariate p-values of association between all possible pairs of SNPs and phenotypes. They have also suggested a hierarchical procedure in which SNPs are first tested by a fixed genome wide level of significance (e.g., Bonferroni correction) and then the B-H rule is applied to detect the non-null traits for a selected SNP using a level of FDR adjusted for the selection in the first step. We employed their strategy in our application to the GERA cohort, but found no SNP to be associated with two or more phenotypes. On the contrary, $MBH_{0.01}$ picked up four significant SNPs associated with two non-null phenotypes that substantially overlap with the known associations from the NHGRI-EBI catalog. Hence their procedure seems to be too conservative as our data comprises more than $50K$ individuals.

Of note, due to the technical adaptability of LASSO in high-dimensional regression, in the context of a gene-based association analysis for a complex trait, A-LASSO can be applied based on the regression (linear or logistic) of the phenotype on the SNP set/gene in order to identify an optimal non-null subset of genetic variants underlying a GW signal of gene-level association. Similarly, the modified B-H procedure can also be employed for this purpose.

Our study presents an explicit evaluation via simulation and application of multiple different strategies for selecting non-null traits underlying a pleiotropic signal. A limitation of our study is that none of the criteria for selecting non-null traits condition on initial testing of overall signals within the same data (discussed above). The implications of this deserve further investigation. When testing for multivariate association, the true subset of associated traits is unknown. Including too many irrelevant phenotypes may lead to loss in power due to increased degrees of freedom of the test. Since the modified B-H procedure performs well, one might first use this to select the most important phenotypes, and then perform a multivariate test of association based on only the selected traits. Such an approach allows for simultaneous inference on the non-null subset of traits and evidence of pleiotropic association like *ASSET*, but with higher specificity and sensitivity of the selected non-null traits. An open problem here is how to adjust for the selection simultaneously in the testing

procedure – computing the p-value of pleiotropic association conditioning on the selection of non-null traits – to avoid potential increase in the false positive rate.

In summary, our simulations indicate that *MultiPhen* is overall more powerful than *ASSET* for assessing pleiotropy. The modified B-H procedure, in particular $MBH_{0.01}$, provides more reliable selection of non-null traits compared to the other criteria considered. This procedure is computationally easy to implement, and its application to existing cohort data detected pleiotropic signals that agree with previously published results. Nevertheless, trait selection approaches that simultaneously model multiple phenotypes and genotype are also promising and merit further development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health grants R01CA088164, U01CA127298, R25CA112355, and the UCSF Goldberg-Benioff Program in Cancer Translational Biology. We thank Thomas Hoffman, Joel Mefford, Samsiddhi Bhattacharya, and Michael Passarelli for important discussions relating to this work. The GERA cohort data came from a grant, the Resource for Genetic Epidemiology Research in Adult Health and Aging (RC2 AG033067; Schaefer and Risch, PIs) awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. The RPGEH was supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, Kaiser Permanente Northern California, and the Kaiser Permanente National and Northern California Community Benefit Programs. The RPGEH and the Resource for Genetic Epidemiology Research in Adult Health and Aging are described in the following publication, Schaefer C, et al., The Kaiser Permanente Research Program on Genes, Environment and Health: Development of a Research Resource in a Multi-Ethnic Health Plan with Electronic Medical Records, In preparation.

Appendix: An alternative way of estimating correlations between β^s

To estimate the correlation matrix among SNP association statistics for different traits, it is sufficient to know how to calculate a non-diagonal element. Thus, we outline how to estimate the correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$ under the null hypothesis of no pleiotropic association. Let y_1, y_2 denote the sampled values of two traits Y_1, Y_2. Suppose Y_j is regressed on X (with sampled value x) and the likelihood of Y_j considered in the regression is denoted by $f(y_j|x, \beta_j)$, $j = 1, 2$.

Let $S_j(\beta_j) = \frac{\partial log f(y_j|x,\beta_j)}{\partial \beta_j}$ be the score function for β_j, and $\hat{\beta}_j$ be its MLE. Let $I_{jj}(\beta_j)$ be Fisher's information for β_j, which is defined as $E[(\frac{\partial log f(y_j|x,\beta_j)}{\partial \beta_j})^2]$. If $\log f(y_j|x, \beta_j)$ is differentiable twice with respect to β_j, then $I_{jj}(\beta_j) = -E[\frac{\partial^2 log f(y_j|x,\beta_j)}{\partial^2 \beta_j}]$. In standard cases, e.g., when the trait is continuous and distributed as normal, or is binary with a logistic distribution, $I_{jj}(\beta_j)$ becomes free of the values of phenotype, and hence is the same as:

$$-\frac{\partial^2 log f(y_j|x,\beta_j)}{\partial^2 \beta_j}$$

Using the Taylor's series expansion of the score function around the true value of $\beta_j = \beta_{j0}$, one can obtain: $S_j(\beta_j) \approx S_j(\beta_{j0}) - I_{jj}(\beta_{j0})(\beta_j - \beta_{j0})$. Since, by the definition of the MLE, $S_j(\beta_j)|_{\beta_j=\hat{\beta}_j}=0$,

$$\begin{aligned} S_j\left(\hat{\beta}_j\right) &\approx S_j(\beta_{j0}) \\ &- I_{jj}(\beta_{j0})\left(\hat{\beta}_j - \beta_{j0}\right) \Rightarrow 0 \approx S_j(\beta_{j0}) \\ &- I_{jj}(\beta_{j0})\left(\hat{\beta}_j - \beta_{j0}\right) \Rightarrow \hat{\beta}_j \approx \beta_{j0} \\ &+ I_{jj}^{-1}(\beta_{j0}) S_j(\beta_{j0}),\ j=1,2 \end{aligned}$$

Under the null hypothesis of no global association: $\beta_{10} = 0$, $\beta_{20} = 0$, and hence,

$$\mathrm{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) \approx I_{11}^{-1}(\beta_{10})\,\mathrm{Cov}\left[S_1(\beta_{10}), S_2(\beta_{20})\right] I_{22}^{-1}(\beta_{20}),\ \text{and}\ \mathrm{Var}\left(\hat{\beta}_j\right) = I_{jj}^{-2}(\beta_{j0})\,\mathrm{Var}\left[S_j(\beta_{j0})\right],\ j=1,2.$$

So, $\mathrm{corln}\left(\hat{\beta}_1, \hat{\beta}_2\right) = \dfrac{\mathrm{Cov}\left[S_1(\beta_{10}), S_2(\beta_{20})\right]}{\sqrt{\mathrm{Var}\left[S_1(\beta_{10})\right]}\sqrt{\mathrm{Var}\left[S_2(\beta_{20})\right]}}$. The variance-covariance matrix for $S_j(\beta_{j0})$, $j=1, 2$, can be estimated based on the values of the score functions calculated at $\beta_j = 0$ for the individuals in the sample. For separate studies of multiple phenotypes, the covariance term can be estimated in similar way only using shared individuals across studies.

References

Akaike, H. Breakthroughs in statistics. Springer; 1992. Information theory and an extension of the maximum likelihood principle; p. 610-624.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological). 1995:289–300.

Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Annals of statistics. 2001:1165–1188.

Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, Yeager M, Chung CC, Chanock SJ, Chatterjee N, et al. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. The American Journal of Human Genetics. 2012; 90(5):821–835. [PubMed: 22560090]

Bower JK, Appel LJ, Matsushita K, Young JH, Alonso A, Brancati FL, Selvin E. Glycated hemoglobin and risk of hypertension in the atherosclerosis risk in communities study. Diabetes care. 2012; 35(5):1031–1037. [PubMed: 22432110]

Chen J, Chen Z. Extended bayesian information criteria for model selection with large model spaces. Biometrika. 2008; 95(3):759–771.

Chen J, Chen Z. Extended bic for small-n-large-p sparse glm. Statistica Sinica. 2012; 22(2):555.

Friedman J, Hastie T, Tibshirani R. glmnet: Lasso and elastic-net regularized generalized linear models. R package version. 2009; 1

Galesloot TE, van Steen K, Kiemeney LA, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. PloS one. 2014; 9(4):e95923. [PubMed: 24763738]

Guo X, Li Y, Ding X, He M, Wang X, Zhang H. Association tests of multiple phenotypes: Atemp. PloS one. 2015; 10(10):e0140348. [PubMed: 26479245]

Klei L, Luca D, Devlin B, Roeder K. Pleiotropy and principal components of heritability combine to increase power for association analysis. Genetic Epidemiology. 2008; 32:9–19. [PubMed: 17922480]

Majumdar, A.; Mukhopadhyay, I.; Ghosh, S. BMC proceedings. Vol. 8. BioMed Central Ltd; 2014. Association mapping of blood pressure levels in a longitudinal framework using binomial regression; p. S74

Majumdar A, Witte JS, Ghosh S. Semiparametric allelic tests for mapping multiple phenotypes: Binomial regression and mahalanobis distance. Genetic epidemiology. 2015; 39(8):635–650. [PubMed: 26493781]

Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nature genetics. 2013; 45(4):353–361. [PubMed: 23535729]

O'Reilly PF, Hoggart CJ, Pomyen Y, et al. Joint model of multiple phenotypes can increase discovery in gwas. PLoS ONE. 2012; 7(5):e34861. [PubMed: 22567092]

Peterson C, Bogomolov M, Benjamini Y, Sabatti C. Many phenotypes without many false discoveries: Error controlling strategies for multi-traits association studies. Genetic Epidemiology. 2016; 40(1): 45–56. [PubMed: 26626037]

Schwarz G, et al. Estimating the dimension of a model. The annals of statistics. 1978; 6(2):461–464.

Stephens M. A unified framework for association analysis with multiple related phenotypes. PloS one. 2013; 8(7):e65245. [PubMed: 23861737]

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

Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996:267–288.

Wang K. Testing genetic association by regressing genotype over multiple phenotypes. PLoS ONE. 2014:e106918. [PubMed: 25221983]

Witte JS. Multiple prostate cancer risk variants on 8q24. Nature genetics. 2007; 39(5):579–580. [PubMed: 17460686]

Wu B, Pankow JS. Statistical methods for association tests of multiple continuous traits in genome-wide association studies. Annals of human genetics. 2015

Yan T, Li Q, Li Y, Li Z, Zheng G. Genetic association with multiple traits in the presence of population stratification. Genetic epidemiology. 2013; 37(6):571–580. [PubMed: 23740720]

Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, Smith JA, Yanek LR, Sun YV, Edwards TL, et al. Meta-analysis of correlated traits via summary statistics from gwass with an application in hypertension. The American Journal of Human Genetics. 2014

Zou H. The adaptive lasso and its oracle properties. Journal of the American statistical association. 2006; 101(476):1418–1429.

Zou H, Hastie T, Tibshirani R, et al. On the degrees of freedom of the lasso. The Annals of Statistics. 2007; 35(5):2173–2192.

Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nature genetics. 2013; 45(11):1274–1283. [PubMed: 24097068]

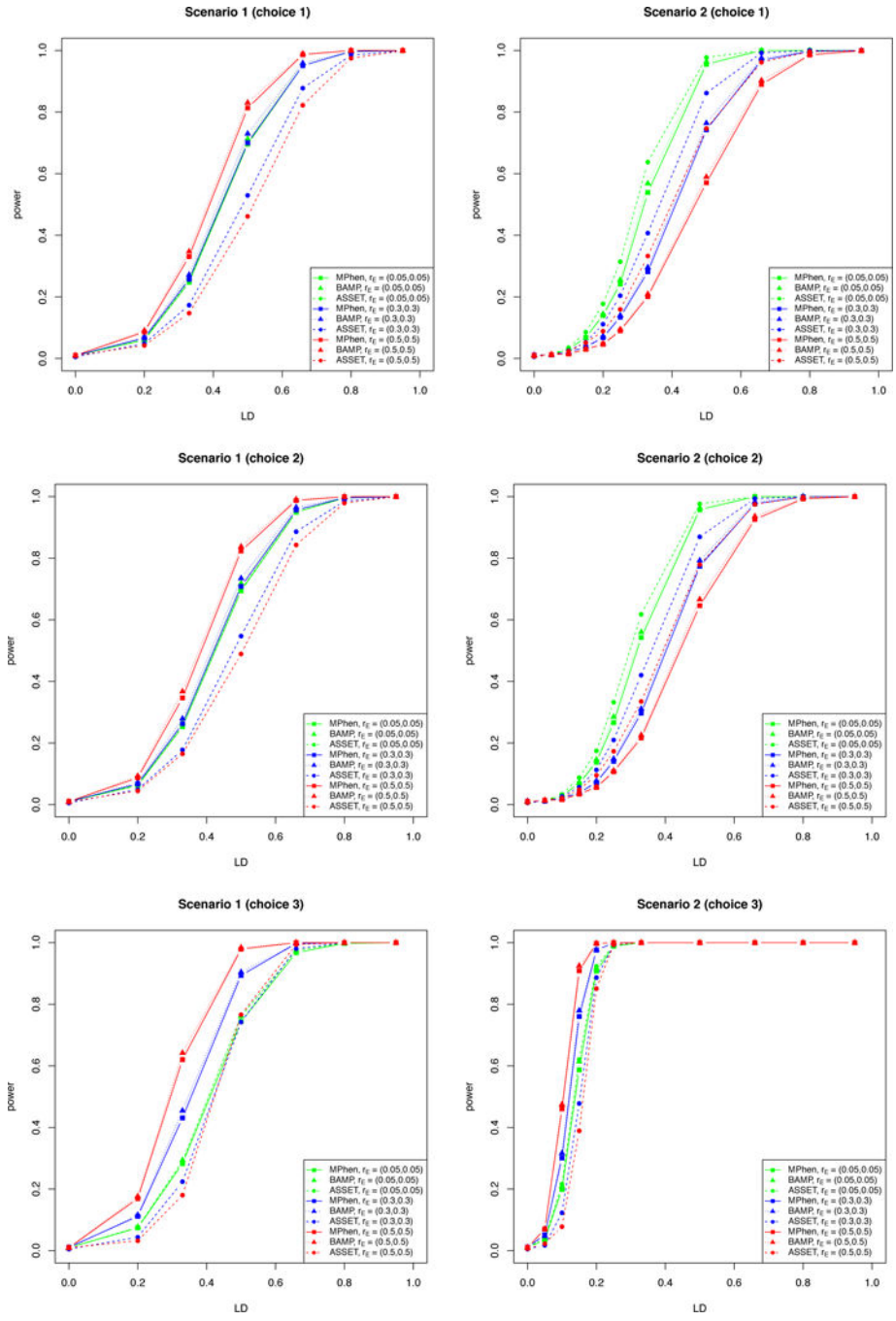

Figure 1.

Power comparison among *MultiPhen*, *BAMP*, and *ASSET* for assessing pleiotropy in simulation Scenario 1 and 2 (Table 1). Different colors represent different degree of phenotypic correlations. r_E denotes the residual correlation structure of the phenotypes.

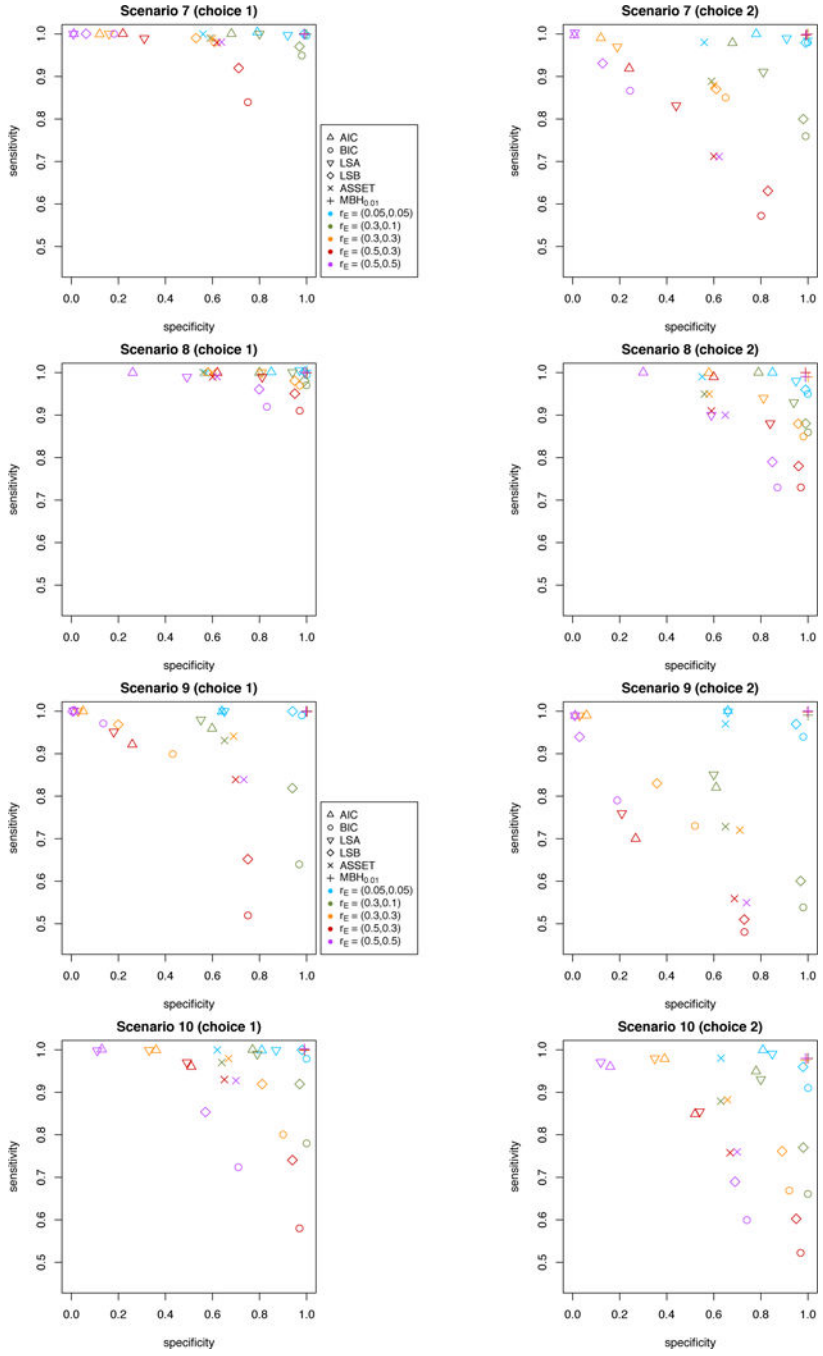

Figure 2.

Sensitivity and specificity of different approaches for determining the subset of traits underlying pleiotropic association in simulation Scenario 7, 8, 9, and 10 (Table 3). Different colors represent different degree of phenotypic correlations. r_E denotes the residual correlation structure of the phenotypes.

Table 1

Description of simulation scenarios in the study comparing power of *BAMP*, *MultiPhen*, and *ASSET* for assessing pleiotropy. K is the total number of traits of which K_1 are associated. h^2 denotes the trait-specific heritabilities for K traits. The direction of QTL effects across traits (βs) is also provided.

Scenario	K_1/K	Results	choices of trait-specific heritabilities and direction of effects sizes
1	2/4	Figure 1	Traits are continuous and normally distributed 1: h^2 = (0.3%, 0.3%, 0, 0); (β_1, $\beta_2 > 0$); ($\beta_1 = \beta_2$) 2: h^2 = (0.2%, 0.4%, 0, 0); (β_1, $\beta_2 > 0$); ($\beta_1 \neq \beta_2$) 3: h^2 = (0.3%, −0.3%, 0, 0); ($\beta_1 > 0$, $\beta_2 < 0$); ($\beta_1 = -\beta_2$)
2	4/4	Figure 1	Traits are continuous and normally distributed 1: h^2 = (0.3%, 0.3%, 0.3%, 0.3%); (β_1, β_2, β_3, $\beta_4 > 0$); ($\beta_1 = \beta_2 = \beta_3 = \beta_4$) 2: h^2 = (0.2%, 0.4%, 0.2%, 0.4%); (β_1, β_2, β_3, $\beta_4 > 0$); ($\beta_1 = \beta_3 \neq \beta_2 = \beta_4$) 3: h^2 = (0.3%, −0.3%, 0.3%, −0.3%); (β_1, $\beta_3 > 0$; β_2, $\beta_4 < 0$); ($\beta_1 = -\beta_2 = \beta_3 = -\beta_4$)
3	2/4	Figure S1	same as in Scenario 1 with all the continuous traits dichotomized to binary traits
4	4/4	Figure S1	same as in Scenario 2 with all the continuous traits dichotomized to binary traits
5	2/4	Figure S2	same as in Scenario 1 with all the continuous traits distributed as chi-square
6	4/4	Figure S2	same as in Scenario 2 with all the continuous traits distributed as chi-square

Table 2

Type I error rates for *BAMP*, *MultiPhen*, and *ASSET* for detecting overall pleiotropy (nominal rate = 0.01). h^2 denotes the choices of trait-specific heritabilities. r_E denotes the residual correlation structure of the phenotypes.

		r_E = (0.05, 0.05)			r_E = (0.3, 0.3)			r_E = (0.5, 0.5)		
Scenario	h^2	***BAMP***	***MultiPhen***	***ASSET***	***BAMP***	***MultiPhen***	***ASSET***	***BAMP***	***MultiPhen***	***ASSET***
1	choice 1	0.0086	0.0096	0.0098	0.0082	0.0098	0.0052	0.0098	0.0102	0.0096
	choice 2	0.009	0.0098	0.0112	0.0084	0.0102	0.0054	0.0096	0.0098	0.009
	choice 3	0.009	0.0094	0.0098	0.0076	0.0108	0.0052	0.0096	0.0094	0.0092
2	choice 1	0.0092	0.01	0.0098	0.0112	0.0108	0.0058	0.01	0.0092	0.0074
	choice 2	0.0086	0.0102	0.0096	0.0112	0.0108	0.0056	0.0104	0.0098	0.007
	choice 3	0.0094	0.0092	0.0112	0.0106	0.0114	0.0052	0.0102	0.01	0.0066
3	choice 1	0.0104	0.0106	0.0132	0.0076	0.0094	0.0076	0.0108	0.0108	0.0076
	choice 2	0.0124	0.0102	0.015	0.0094	0.0098	0.007	0.01	0.0088	0.008
	choice 3	0.0118	0.0122	0.0132	0.0092	0.0102	0.008	0.0124	0.0132	0.0076
4	choice 1	0.01	0.0102	0.0132	0.008	0.0072	0.0056	0.0108	0.0092	0.0062
	choice 2	0.009	0.0094	0.0106	0.0086	0.009	0.0088	0.0086	0.0094	0.009
	choice 3	0.0116	0.0114	0.0142	0.0112	0.009	0.009	0.0106	0.0092	0.0074
5	choice 1	0.0106	0.01	0.014	0.01	0.0094	0.0096	0.0072	0.008	0.0062
	choice 2	0.0108	0.01	0.014	0.0102	0.0092	0.0098	0.0072	0.008	0.0062
	choice 3	0.011	0.01	0.0138	0.0098	0.009	0.009	0.0084	0.009	0.0064
6	choice 1	0.0106	0.0096	0.0136	0.0112	0.0092	0.0092	0.0104	0.0114	0.008
	choice 2	0.0108	0.0096	0.0134	0.0112	0.0092	0.009	0.0104	0.0112	0.008
	choice 3	0.0104	0.0104	0.014	0.0098	0.0084	0.0086	0.0106	0.0116	0.0092

Table 3

Description of the simulation scenarios for evaluating selection performance across the different methods. K is the total number of traits of which K_1 are associated. h^2 denotes the trait-specific heritabilities for K traits. The direction of QTL effects across traits (βs) is also provided.

Scenario	K_1/K	Results	choices of the trait-specific heritabilities and direction of the QTL effects
7	2/4	Table S1/Figure 2	Traits are continuous and normally distributed. 1: h^2 = (0.3%, 0.3%, 0, 0); (β_1, $\beta_2 > 0$); ($\beta_1 = \beta_2$) 2: h^2 = (0.2%, 0.4%, 0, 0); (β_1, $\beta_2 > 0$); ($\beta_1 \neq \beta_2$) 3: h^2 = (0.3%, −0.3%, 0, 0); ($\beta_1 > 0$, $\beta_2 < 0$); ($\beta_1 = -\beta_2$)
8	2/4	Table S2/Figure 2	same as in Scenario 7 with all the continuous traits dichotomized to binary traits
9	4/8	Table S3/Figure 2	Traits are continuous and normally distributed. 1: h^2 = (0.3%, 0.3%, 0.3%, 0.3%, 0, 0, 0, 0); (β_1, β_2, β_3, $\beta_4 > 0$); ($\beta_1 = \beta_2 = \beta_3 = \beta_4$) 2: h^2 = (0.2%, 0.4%, 0.2%, 0.4%, 0, 0, 0, 0); (β_1, β_2, β_3, $\beta_4 > 0$); ($\beta_1 = \beta_3 \neq \beta_2 = \beta_4$) 3: h^2 = (0.3%, −0.3%, 0.3%, −0.3%, 0, 0, 0, 0); (β_1, $\beta_3 > 0$; β_2, $\beta_4 < 0$); ($\beta_1 = -\beta_2 = \beta_3 = -\beta_4$)
10	4/8	Table S4/Figure 2	same as in Scenario 9 with all the continuous traits dichotomized to binary traits
11	4/8	Table S5	same as in Scenario 9 with all the continuous traits distributed as chi-square
12	3/8	Table S6	Traits are continuous and normally distributed. 1: h^2 = (0.3%, 0.3%, 0.3%, 0, 0, 0, 0, 0), (β_1, β_2, $\beta_3 > 0$); ($\beta_1 = \beta_2 = \beta_3$) 2: h^2 = (0.2%, 0.4%, 0.2%, 0, 0, 0, 0, 0); (β_1, β_2, $\beta_3 > 0$); ($\beta_1 = \beta_3 \neq \beta_2$) 3: h^2 = (0.3%, −0.3%, 0.3%, 0, 0, 0, 0, 0); (β_1, $\beta_3 > 0$; $\beta_2 < 0$); ($\beta_1 = -\beta_2 = \beta_3$)
13	6/8	Table S7	Traits are continuous and normally distributed. 1: h^2 = (0.3%, 0.3%, 0.3%, 0.3%, 0.3%, 0.3%, 0, 0), (β_1, β_2, β_3, β_4, β_5, $\beta_6 > 0$); (β_1=β_2=β_3=β_4=β_5=β_6) 2: h^2 = (0.2%, 0.4%, 0.2%, 0.4%, 0.2%, 0.4%, 0, 0), (β_1, β_2, β_3, β_4, β_5, $\beta_6 > 0$); (β_1=β_3=$\beta_5 \neq$ β_2=β_4=β_6) 3: h^2 = (0.3%,−0.3%, 0.3%, −0.3%, 0.3%, −0.3%, 0, 0), (β_1, β_3, $\beta_5 > 0$; β_2, β_4, $\beta_6 < 0$); ($\beta_1 = -\beta_2 = \beta_3$ $= -\beta_4 = \beta_5 = -\beta_6$)
14	8/20	Table S8	Traits are continuous and normally distributed. 1: h^2 = (0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0); (β_1, β_2, β_3, β_4, β_5, β_6, β_7, $\beta_8 > 0$); ($\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8$); 2: h^2 = (0.002, 0.004, 0.002, 0.004, 0.002, 0.004, 0.002, 0.004, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) (β_1, β_2, β_3, β_4, β_5, β_6, β_7, $\beta_8 > 0$); ($\beta_1 = \beta_3 = \beta_5 = \beta_7 \neq \beta_2 = \beta_4 = \beta_6 = \beta_8$); 3: h^2 = (0.003, −0.003, 0.003, −0.003, 0.003, −0.003, 0.003, −0.003, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0); (β_1, β_3, β_5, $\beta_7 > 0$; β_2, β_4, β_6, $\beta_8 < 0$);($\beta_1 = -\beta_2 = \beta_3 = -\beta_4 = \beta_5 = -\beta_6 = \beta_7 = -\beta_8$)
15–18		Table S9, S10, S11, S12	Simulations in Table S1, S2, S3, S4 are repeated changing the MAF at the marker locus from 0.1 to 0.2.

Table 4

Correlation among the selected four phenotypes in GERA cohort.

	HYP	T2D	CVD	CAN
HYP	1.00	–	–	–
T2D	0.28	1.00	–	–
CVD	0.33	0.17	1.00	–
CAN	0.15	0.05	0.16	1.00

Genet Epidemiol. Author manuscript; available in PMC 2016 July 01.

Table 5

Genome wide significant SNPs associated with multiple phenotypes in the GERA cohort (p-value < 5×10^{-8}). SNPs identified by both of the methods are colored green; the signals which are detected by *ASSET* but not by *MultiPhen* are colored red; and the signals which are picked up by *MultiPhen* but not by *ASSET* are colored blue. Those not detected at genome wide level of significance were generally just beyond the cut-off.

			Pleiotropic association p-values			Phenotypes associated with mapped genes	
Chr	***rs*ID**	**HWEpv**	***ASSET***	***MultiPhen***	**Mapped genes***	**UCSC genome browser**	**NHGRI-EBI GWAS catalog**
1	*rs*1408420	0.51	3.33×10^{-8}	3.64×10^{-8}	RCC2	Basal cell carcinoma, Hepatitis C	Basal cell carcinoma
2	*rs*1275988	0.81	2.36×10^{-10}	6.07×10^{-10}	KCNK3 (1.3KB)	Anorexia nervosa, Insulin	
2	*rs*3769823	0.11	2.76×10^{-8}	4.34×10^{-9}	CASP8	Cancers, Diabetes, Bone mineral density	Cancers, Rheumatoid arthritis
3	*rs*11709077	0.58	1.61×10^{-8}	5.58×10^{-8}			
3	*rs*1470580	0.11	1.33×10^{-8}	7.22×10^{-9}			
5	*rs*183671	0.05	3.64×10^{-8}	1.12×10^{-7}			
6	*rs*12203592	0.007	9.21×10^{-48}	1.52×10^{-45}	IRF4	Leukemia, Celiac disease, Hair color, Freckling, Tanning phenotypes	Sunburns, Facial pigmentation color, Non-melanoma skin cancer, Hair color
6	*rs*872071	0.8	3.41×10^{-8}	7.12×10^{-8}			
7	*rs*849135	0.1	2.88×10^{-12}	2.66×10^{-12}			
8	*rs*13266634	0.97	1.53×10^{-8}	5.59×10^{-8}	SLC30A8	T2D, Glycated hemoglobin levels, Coronary Disease, Myocardial Infarction, Diabetes, Asthma, Obesity, Insulin resistance, Prostate cancer	Intraocular pressure, Proinsulin levels, Asthma, Glycated hemoglobin levels, Fasting glucose-related traits
9	*rs*10116277	0.14	3.81×10^{-8}	1.49×10^{-8}			
10	*rs*12255678	0.88	8.28×10^{-10}	7.58×10^{-10}	TCF7L2	T2D, Diabetes, Coronary Disease	T2D
10	*rs*7079711	0.77	3.47×10^{-10}	1.24×10^{-10}	TCF7L2	Fasting glucose-related traits, Hemoglobin A	Glycated hemoglobin levels
10	*rs*4506565	0.95	3.57×10^{-52}	4.8×10^{-49}	TCF7L2	Two-hour glucose challenge, Cancers	Proinsulin levels
10	*rs*7896811	0.005	1.91×10^{-7}	4.35×10^{-8}	TCF7L2	Obesity, Blood Pressure Determination	Two-hour glucose challenge
11	*rs*67279079	0.49	6.21×10^{-9}	2.78×10^{-7}	TYR (17.6KB)	Multiple sclerosis, Melanoma, Cancers, Tanning phenotypes, Eye colors, Vitiligo	Sunburns, Skin sensitivity to sun, Freckles, Skin pigmentation, Eye color, Vitiligo
16	*rs*12922197	0.006	6.04×10^{-12}	3.48×10^{-12}	CDK10 (2.3KB)	Heart Failure, Hair Color, Melanoma	Melanoma, Hair color, Height
					SPATA33 (7.1KB)		Facial pigmentation, Height, Psychiatric disorders
					SPATA33 (7.1KB)		
					SPATA2L (17.9KB)		
16	*rs*4408545	0.022	1.60×10^{-7}	4.35×10^{-8}	CENPBD1 (5.1KB)		

			Pleiotropic association p-values			Phenotypes associated with mapped genes	
Chr	***rs*ID**	**HWEpv**	***ASSET***	***MultiPhen***	**Mapped genes***	**UCSC genome browser**	**NHGRI-EBI GWAS catalog**
20	*rs*4911442	0.19	5.93×10^{-11}	1.38×10^{-9}	DEF8 (9.6KB) NCOA6	Birth Weight, Diabetes Mellitus, Lung cancer, Tobacco Use Disorder, Plasma HDL cholesterol levels	

*If a SNP maps to a gene but falls outside its boundaries, the SNP's distance from the nearest of the two boundaries of the gene is provided beside the name of the gene.

Table 6

Key aspects of the different selection criteria considered here.

Selection criteria	Explores all subsets	Regression type	Multivariate or univariate	Brief description
AIC	Yes	Genotype on phenotypes	Multivariate	Model selection criterion
BIC	Yes	Genotype on phenotypes	Multivariate	Model selection criterion
EBIC	Yes	Genotype on phenotypes	Multivariate	Model selection criterion
LSA	No	Genotype on phenotypes	Multivariate	Adaptive LASSO + *AIC*
LSB	No	Genotype on phenotypes	Multivariate	Adaptive LASSO + *BIC*
ASSET	Yes	Phenotypes on genotype	Univariate	Subset-based fixed effects meta analysis
MBH_w	No	Phenotypes on genotype	Univariate	B-H procedure with little modification

Table 7

The selection of non-null traits for the genome wide significant SNPs associated with multiple traits in the GERA cohort where at least two of the criteria detected multiple traits. Green color denotes the selection of a trait. Red color represents the exclusion of a trait from the subset of non-null traits. The univariate results are denoted by *UNIV* in which green denotes inclusion of the trait at p-value $< 5 \times 10^{-8}$, yellow indicates $5 \times 10^{-8} <$ p-value < 0.05, and red denotes p-value > 0.05.

SNP	Disease*	*UNIV*	*AIC*	*LSA*	*BIC*	*LSB*	*ASSET*	$MBH_{0.05}$	$MBH_{0.01}$
chr3:*rs*1470580	HYP								
	T2D ✓								
	CVD								
	CAN								
chr6:*rs*12203592	HYP								
	T2D								
	CVD								
	CAN ✓								
chr6:*rs*872071	HYP								
	T2D								
	CVD								
	CAN ✓								
chr8:*rs*13266634	HYP								
	T2D ✓								
	CVD								
	CAN								
chr9:*rs*10116277	HYP								
	T2D ✓								
	CVD ✓								
	CAN ✓								
chr10:*rs*12255678	HYP ✓								
	T2D ✓								
	CVD								
	CAN ✓								

SNP	Disease*	*UNIV*	*AIC*	*LSA*	*BIC*	*LSB*	*ASSET*	$MBH_{0.05}$	$MBH_{0.01}$
chr10:*rs*7079711	HYP								
	T2D								
	CVD								
	CAN								
chr10:*rs*4506565	HYP ✓								
	T2D ✓								
	CVD								
	CAN ✓								
chr11:*rs*67279079	HYP								
	T2D								
	CVD								
	CAN ✓								

*The checkmark (✓) symbol beside a disease indicates that the corresponding SNP was found to be in LD with at least one SNP ($r^2 \geq 0.2$ in European populations) which has been reported to be associated with the disease (or a closely related phenotype) in the NHGRI-EBI GWAS catalog (described in Tables S15 and S16 in the supplementary material).