

# UCSF

## UC San Francisco Previously Published Works

### Title

Semiparametric Allelic Tests for Mapping Multiple Phenotypes: Binomial Regression and Mahalanobis Distance

### Permalink

<https://escholarship.org/uc/item/2q84k3xp>

### Journal

Genetic Epidemiology, 39(8)

### ISSN

0741-0395

### Authors

Majumdar, Arunabha  
Witte, John S  
Ghosh, Saurabh

### Publication Date

2015-12-01

### DOI

10.1002/gepi.21930

Peer reviewed



Published in final edited form as:

*Genet Epidemiol.* 2015 December ; 39(8): 635–650. doi:10.1002/gepi.21930.

## Semi-parametric Allelic Tests For Mapping Multiple Phenotypes: Binomial Regression And Mahalanobis Distance

Arunabha Majumdar<sup>1,2,\*</sup>, John S. Witte<sup>1</sup>, and Saurabh Ghosh<sup>2</sup>

<sup>1</sup>University of California, San Francisco

<sup>2</sup>Indian Statistical Institute, Kolkata

### Abstract

Binary phenotypes commonly arise due to multiple underlying quantitative precursors. Genetic variants may impact multiple traits in a pleiotropic manner. Hence, simultaneously analyzing such correlated traits may be more powerful than analyzing individual traits. Various genotype-level methods, e.g. *MultiPhen* [O'Reilly et al., 2012], have been developed to identify genetic factors underlying a multivariate phenotype. For univariate phenotypes, the usefulness and applicability of allele-level tests have been investigated. The test of allele frequency difference among cases and controls is commonly used for mapping case-control association. However, allelic methods for multivariate association mapping have not been studied much. We explore two allelic tests of multivariate association: one using a Binomial regression model based on inverted regression of genotype on phenotype (*BAMP*), and the other employing the Mahalanobis distance between two sample means of the multivariate phenotype vector for two alleles at a SNP (*DAMP*). These methods can incorporate both discrete and continuous phenotypes. Some theoretical properties for *BAMP* are studied. Using simulations, the power of the methods for detecting multivariate association are compared with the genotype-level test *MultiPhen*. The allelic tests yield marginally higher power than *MultiPhen* for multivariate phenotypes. For one/two binary traits under recessive mode of inheritance, allelic tests are found substantially more powerful. All three tests are applied to two real data and the results offer some support for the simulation study. Since the allelic approaches assume Hardy-Weinberg Equilibrium (HWE), we propose a hybrid approach for testing multivariate association that implements *MultiPhen* when HWE is violated and *BAMP* otherwise.

### Keywords

Genome-wide association study; multiple phenotypes; multivariate association; allelic tests

### Introduction

A clinical trait or disease with a complex end-point is generally characterized by multiple quantitative precursors. In such situations, one can improve statistical power by considering

\* Address for correspondence: Arunabha Majumdar, Department of Epidemiology and Biostatistics, University of California, San Francisco, 1450 3<sup>rd</sup> St, Box 3110, Witte Lab, San Francisco, CA 94158, Arunabha.Majumdar@ucsf.edu.

The authors do not have any conflicts of interest.

a multivariate phenotype [Lange et al., 2004; Klei et al., 2008; O'Reilly et al., 2012; Stephens, 2013; Zhu et al., 2014]. Moreover, in the presence of association with multiple traits, a multivariate approach is biologically more consistent than a univariate approach. Galesloot et al. [2014] showed that the use of multivariate association analyses can yield higher power compared to univariate analyses even when genetic correlations between traits are weak or when a genetic variant is only associated with one of the phenotypes (that is, when there is no multivariate association). For example, cardiovascular disease (CVD) is characterized by high levels of both low-density serum lipoprotein levels (LDL) and systolic blood pressure (SBP) [Newman et al., 1986]. In order to identify genomic regions associated with CVD, it may be intuitively more appealing and statistically more powerful to carry out association tests with a bivariate phenotype vector comprising LDL and SBP instead of the binary CVD phenotype.

To analyze such data, one can implement a multivariate regression approach such as MANOVA, but this requires that the vector of traits have an approximate multivariate normal distribution. This assumption may be difficult to satisfy for a vector comprising mixed phenotypes, such as, when one trait is binary and the other is continuous. Several methods using variations of generalized estimating equations (GEE) have been developed to incorporate non-normal traits [Zeger and Liang, 1986; Liu et al., 2009]. A log-linear approach [Lee et al., 2011] and a Bayesian network approach [Hartley et al., 2012] have been used to model multiple categorical phenotypes. The linear mixed model approach [Korte et al., 2012] provides a flexible framework to account for the correlation structure between SNPs as well as between individuals in a genome-wide association scan of multiple phenotypes. Other existing methods such as likelihood-based variance components [Almasy and Blangero, 2010] are susceptible to violation in underlying distributional assumptions. Another approach that circumvents the problem of modeling the multivariate phenotype is to obtain a reduced univariate phenotype using principal components [Klei et al., 2008; Minica et al., 2010]. However, association results based on principal components may be difficult to interpret biologically. For example, the first principal component of LDL and SBP is a linear combination of the two traits that explains the highest variability among the traits. However, it is difficult to understand the biological meaning or genetic underpinnings of such a derived trait.

An alternative approach to association mapping of multivariate phenotypes is to reverse the conventional model and regress genotypes on the multivariate phenotype. One such approach has been termed *MultiPhen* [O'Reilly et al., 2012] and has been shown to yield high power for detecting association with multiple traits [Galesloot et al., 2014]. The advantages of such an approach include the flexibility of incorporating traits of mixed type (binary, discrete or continuous) in the vector of phenotypes, allowing for heterogeneity in both the extent and the direction of association between the genetic variant and each constituent phenotype, robustness to possible violations in underlying distributional assumptions and the possibility of adjustment for covariates simultaneously with the multiple phenotypes.

For univariate traits, allele-level test for association has been investigated comparing its performance with genotype-level test [Guedj et al., 2008; Zheng, 2008; Lee et al., 2013]. To

the best of our knowledge, careful studies of allele-level tests for mapping association with multiple traits have not been undertaken. In this article, we explore two allelic tests for association with multiple phenotypes: one based on a binomial regression model (*BAMP*, Binomial regression-based Association of Multivariate Phenotypes) and the other an extension of the allelic test for univariate traits proposed by Lee et al. [2013] to include multiple phenotypes (*DAMP*, Distance-based Association of Multivariate Phenotypes). We can view *BAMP* as a suitable modification of *MultiPhen* [O'Reilly et al., 2012] at the allelic level and show that it is equivalent to the classical case control allele-based test of association for binary traits. *BAMP* is proposed in the framework of inverted regression of genotype on phenotypes whereas *DAMP* is formulated as a test of difference between the two sample means of vectors of the multivariate phenotype corresponding to the two alleles at a SNP under the usual set-up of regressing phenotype on genotype.

Using simulations under a wide spectrum of multivariate phenotype models, we compare the performance of *BAMP* and *DAMP* with *MultiPhen*. We find that, for multivariate phenotypes, both the allelic tests produce comparable powers, but marginally higher than *MultiPhen*, in particular, *BAMP* yields a consistent marginal increase of power over *MultiPhen*. It is also found that, for a single or two binary traits under a recessive mode of inheritance, the allelic tests offer substantially higher power than *MultiPhen*. We also demonstrate the advantage of using allelic tests by carrying out association analyses on two real data of multivariate phenotypes: one comprising homocysteine levels, Vitamin B<sub>12</sub> levels and coronary artery disease (CAD) status in a North Indian population, and another comprising four case-control phenotypes - hypertension, diabetes type II, cardiovascular disease, and cancers for European-American population in “Resource for Genet Epidemiol Research on Adult Health and Aging” (GERA) cohort data obtained from dbGaP [dbGaP Study Accession: phs000674.v1.p1]. Based on our observations, we propose a hybrid approach for mapping association with multiple traits that uses *MultiPhen* when variants fail HWE and *BAMP* otherwise.

## Methods

Suppose  $Y = (Y_1, \dots, Y_k)$  denotes a multivariate phenotype vector, where  $Y_j, j = 1, 2, \dots, k$ , can be either a qualitative or a quantitative trait. We assume that the multivariate phenotype is controlled by a biallelic trait locus with alleles  $A_1$  (minor) and  $A_2$ . Consider a biallelic marker locus with alleles  $M_1$  (minor) and  $M_2$  such that the coefficient of linkage disequilibrium between the trait locus and the marker locus is  $\delta$ , which is defined as  $\delta = P(A_1 M_1) - P(A_1)P(M_1)$ . Suppose  $X$  denotes the number of minor alleles (0, 1, or 2) at the marker locus. We denote the frequencies of the alleles  $A_1$  and  $M_1$  by  $p$  and  $m$ , respectively. We assume that data are available on the genotypes at the marker locus and on the  $k$ -dimensional multivariate phenotype vectors for  $n$  unrelated individuals randomly selected from a population. Suppose  $Y_i = (Y_{i1}, \dots, Y_{ik})$  and  $X_i$  represent the multivariate phenotype data and the count of the allele  $M_1$ , respectively, for the  $i^{\text{th}}$  individual,  $i = 1, \dots, n$ .

## MultiPhen

The genotype-level test implemented in *MultiPhen*, is based on a proportional odds logistic regression model [O'Reilly et al., 2012] given by:

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

Thus, the conditional distribution of the genotypes given  $Y = y$  can be obtained as follows:

$$P(X=0|y) = \frac{1}{1 + \exp(\alpha_0 + \sum_{j=1}^k \gamma_j y_j)}, \quad P(X=2|y) = \frac{\exp(\alpha_1 + \sum_{j=1}^k \gamma_j y_j)}{1 + \exp(\alpha_1 + \sum_{j=1}^k \gamma_j y_j)},$$

$$P(X=1|y) = \frac{1}{1 + \exp(\alpha_1 + \sum_{j=1}^k \gamma_j y_j)} - \frac{1}{1 + \exp(\alpha_0 + \sum_{j=1}^k \gamma_j y_j)}.$$

with the restriction that  $\alpha_0 \geq \alpha_1$  (in order to satisfy that  $P(X=1|y)$  is non-negative). A likelihood ratio test can be performed in this proportional odds logistic regression framework corresponding to  $H_0: \bigcap_{j=1}^k \{\gamma_j = 0\}$  ( $\gamma_1 = \dots = \gamma_k = 0$ ) versus  $H_1: \bigcup_{j=1}^k \{\gamma_j \neq 0\}$  (at least one  $\gamma_j \neq 0$ ,  $j=1, \dots, k$ ).

## BAMP

We model the conditional distribution of  $X$  (the number of minor alleles at the marker locus) given  $Y = y$  (the vector of multivariate phenotype) as binomial with parameters 2 and  $p(y)$ , where  $p(y)$  is a logistic link function given by:

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

It follows that  $p(y) = P(M_1|y)$ . The model implicitly assumes that, conditioned on the vector of multivariate phenotype, the marker locus is in Hardy-Weinberg Equilibrium (HWE). The test for association is equivalent to testing  $H_0: \beta_1 = \dots = \beta_k = 0$  versus  $H_1$ : at least one  $\beta_j \neq 0$ ,  $j = 1, \dots, k$ . Under  $H_0$ , the minor allele frequency at the marker locus is independent of  $y$  and is given by:  $P(M_1|y) = \frac{\exp(\alpha)}{1 + \exp(\alpha)}$ . The test can be performed using a likelihood ratio statistic which is distributed as chi-square with  $k$  degrees of freedom under  $H_0$ . We note that this test procedure can be viewed as an allelic modification of *MultiPhen*.

We highlight the following two theoretical properties of *BAMP*:

**Property 1**—Testing for  $H_0: \beta_1 = \dots = \beta_k = 0$  vs  $H_1$ : at least one  $\beta_j \neq 0$ ,  $j = 1, \dots, k$ , is equivalent to testing  $H_0: \delta = 0$  (no association between the marker locus and the traits) vs  $H_1: \delta \neq 0$  (presence of association).

**Property 2**—If the phenotype is binary (in a case-control framework), *BAMP* is equivalent to the usual test of difference in allele frequencies among cases and controls.

The proof of each of the properties is provided in the Appendix. In genome-wide association studies (GWAS), the causal SNPs themselves may not be genotyped, rather the SNPs that are in linkage dis-equilibrium (LD) with them are identified through tests for association. Hence, a detected association between a marker SNP and the phenotypes under consideration is attributed to the non-zero LD between the marker and the causal SNPs. Thus, it is of interest to mathematically prove that testing the null hypothesis of no genotype-phenotype association against the alternative hypothesis of presence of association, is equivalent to testing  $\delta = 0$  versus  $\delta \neq 0$ .

We note that the proof of Property 1 is based on the assumption that the association between a marker SNP and multivariate phenotypes is induced due to non-zero LD ( $\delta \neq 0$ ) between the marker SNP and a single causal SNP instead of multiple independent causal SNPs present in nearby genomic regions. This assumption is crucial for the ease of mathematical derivation underlying the proof. Because, the proof is based on the explicit expression of the joint probability of genotypes at one marker and one causal SNP that are in LD. But, the expression of this joint probability for genotypes at the marker and multiple causal SNPs, with each of which the marker is in LD, can be algebraically intractable.

While proving Property 1,  $\delta$  is used as the definition of LD since it is sufficient for the derivation. Other commonly used formulations of LD, for example,  $r^2$  ( $\delta^2$  divided by the variance of allele frequencies at the two loci) and  $D'$  ( $\delta$  standardized by its minimum and maximum), are essentially derived based on  $\delta$ .

## DAMP

Motivated by the allelic version of the genotype-level test corresponding to a quantitative phenotype [Lee et al., 2013], we propose a suitable multivariate extension as follows. Suppose  $(Y_1, Y_2, \dots, Y_n)$  denote the multivariate phenotype vectors corresponding to  $n$  randomly selected individuals in the population. We consider two groups of observations corresponding to the two alleles  $M_1$  (minor) and  $M_2$  at the marker locus. If the genotype of the  $i^{\text{th}}$  individual is  $M_1M_1$ , we assign two observations each equal to  $Y_i$  in the first group, if the genotype is  $M_2M_2$ , we assign two observations each equal to  $Y_i$  in the second group; and if the genotype is  $M_1M_2$ , we assign one observation equal to  $Y_i$  in each of the two groups. Hence, if the first group is of size  $n_1$  and the second group is of size  $n_2$ , the total number of observations in the two groups is  $2n$ . If the population means of the multivariate phenotype vector  $Y$  conditioned on the two groups are  $\mu_1$  and  $\mu_2$ , respectively, the test for association can be performed by testing  $H_0 : \mu_1 = \mu_2$  versus  $H_1 : \mu_1 \neq \mu_2$ . Assuming that the covariance structure of  $Y(\Sigma)$  is identical in the two groups, for sufficiently large  $n_1$  and  $n_2$ ,

$$\bar{Y}_{M_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_{iM_1} \rightarrow N_k \left( \mu_1, \frac{\Sigma}{n_1} \right), \quad \text{and,} \quad \bar{Y}_{M_2} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_{iM_2} \rightarrow N_k \left( \mu_2, \frac{\Sigma}{n_2} \right).$$

The above asymptotic distributions are obtained using the multivariate central limit theorem (MCLT). The test statistic is based on the Mahalanobis distance between the sample mean vectors of the two groups ( $\bar{Y}_{M_1}$  and  $\bar{Y}_{M_2}$ ). If the sample sizes of the two groups ( $n_1$  and  $n_2$ ) are sufficiently large, then it follows from the MCLT that, the test statistic  $(\bar{Y}_{M_1} - \bar{Y}_{M_2})'S^{-1}(\bar{Y}_{M_1} - \bar{Y}_{M_2})$  is asymptotically distributed as chi-square with  $k$  degrees of freedom, where  $S$  is the combined sample covariance matrix of  $Y$  based on the two groups, and hence a consistent estimator of  $\Sigma$ .

For a binary phenotype, the allelic test based on allele frequencies is asymptotically equivalent to the genotype-based trend test under the assumption of HWE [Guedj et al., 2006, 2008; Sasieni, 1997; Zheng, 2008]. We note that the allele-based method proposed by Lee et al. [2013] in the context of a single quantitative trait uses a linear regression of the phenotype on each of the alleles constituting the genotype of an individual. The above method (*DAMP*) is equivalent to a direct multivariate extension of regressing the vector of phenotypes on the alleles. However, Lee et al. [2013] showed that, considering the allele status (coded as 0 or 1) as a predictor in the linear regression and ignoring the correlation induced by replicating the phenotype data twice, yields marginally lower power compared to using the genotype status (coded as 0, 1 or 2) as a predictor.

## Simulation study

We carried out simulation studies to compare the false positive rates and powers of *MultiPhen*, *BAMP*, and *DAMP* for detecting association with multiple traits.

### Simulation models

We consider a vector of quantitative traits  $Y = (Y_1, \dots, Y_k)$  that is modulated by a biallelic QTL with alleles  $A_1$  and  $A_2$ , and a random vector of environmental components  $e = (e_1, \dots, e_k)$  having mean vector zero and covariance matrix  $\Sigma$ . We assume that  $E(Y_j|A_1A_1) = a_j$ ,  $E(Y_j|A_1A_2) = b_j$ , and  $E(Y_j|A_2A_2) = -a_j$ ,  $i = 1, \dots, k$ . We denote the vectors  $(a_1, \dots, a_k)$  and  $(b_1, \dots, b_k)$  by  $a$  and  $b$ , respectively. We assume two possible different distributions of  $e$ : multivariate normal and chi-square. We induce different degrees of genetic effects on the traits by varying the choices of  $a$  and  $b$ . In order to simulate a vector of phenotypes comprising both binary and quantitative traits, we first generate an observation  $Y$  according to the above model and then dichotomize some of the components by considering different thresholds so as to fix the prevalence of the binary traits at desired levels.

We also consider a biallelic marker locus with alleles  $M_1$  and  $M_2$  which is in linkage disequilibrium (LD) with the QTL. The extent of LD is varied to produce different degrees of associations between the marker locus and the vector of phenotypes. The genetic correlation between a pair of traits, say  $Y_1$  and  $Y_2$ , induced by the QTL, is equal to the correlation between  $E(Y_1|G)$  and  $E(Y_2|G)$ , where  $G$  is the genotype at the QTL and is given by:

$$\frac{\{a_1 a_2 (p^2 + q^2) + 2b_1 b_2 pq\} - \{a_1 (p^2 - q^2) + 2b_1 pq\} \{a_2 (p^2 - q^2) + 2b_2 pq\}}{\sqrt{\{a_1^2 (p^2 + q^2) + 2b_1^2 pq\} - \{a_1 (p^2 - q^2) + 2b_1 pq\}^2} \sqrt{\{a_2^2 (p^2 + q^2) + 2b_2^2 pq\} - \{a_2 (p^2 - q^2) + 2b_2 pq\}^2}}$$

Combining the genetic correlation between  $(Y_1, Y_2)$  with the correlation  $\rho_e$  between the environmental components modulating  $(Y_1, Y_2)$ , the overall correlation induced between  $Y_1$  and  $Y_2$  can be derived as:

$$\frac{\{a_1 a_2 (p^2 + q^2) + 2b_1 b_2 pq + \rho_e \sigma_1 \sigma_2\} - \{a_1 (p^2 - q^2) + 2b_1 pq\} \{a_2 (p^2 - q^2) + 2b_2 pq\}}{\sqrt{\{\sigma_1^2 + a_1^2 (p^2 + q^2) + 2b_1^2 pq\} - \{a_1 (p^2 - q^2) + 2b_1 pq\}^2} \sqrt{\{\sigma_2^2 + a_2^2 (p^2 + q^2) + 2b_2^2 pq\} - \{a_2 (p^2 - q^2) + 2b_2 pq\}^2}}$$

Based on these expressions, for a given choice of  $a$  and  $b$  vectors and other parameters, we can choose  $\rho_e$  accordingly to fix the final correlation between a pair of traits at a fixed level. Different choices of QTL effects considered in each table are described as footnotes.

## Simulation design

We carry out the comparison of *MultiPhen*, *BAMP*, and *DAMP* with respect to the false positive rate and power in testing for association with multiple traits. The type I error rates and powers are calculated based on 2000 individuals and 10,000 replications. We consider the allele frequency at the QTL to be 0.1. Four different choices of the standardized LD between the marker locus and the QTL are chosen as 0.0, 0.33, 0.66, 0.95. HWE is assumed while generating the genotype data.

### Four traits

We first consider a multivariate phenotype that comprises four traits. Since the correlation between the traits may impact power, we consider the overall correlation between a pair of traits at four different levels: 0.0, 0.2, 0.5, 0.8, when the traits are distributed as multivariate normal, and 0.05, 0.2, 0.5, 0.8 when they are distributed as chi-square.

In Table I, we describe the choices of trait specific QTL effects and heritabilities considered in different tables. Initially we assume that the QTL has genetic effects only on the first and the second traits in tables X-XIII. In choices 1, 3, and 4, the heritabilities of the first two traits due to the QTL are fixed at 0.5%. In choice 2, we consider heterogeneity in QTL effects by fixing the genetic heritabilities of the first two non-null traits due to the QTL at two different levels: 0.5% and 1%, respectively. For choice 4, the QTL effects on the two traits are considered to be oppositely directed.

While simulating a vector of phenotypes comprising a binary phenotype, we define an individual to be a case if  $Y_1 > C$ , and a control, otherwise. We replace  $Y_1$  by the case or control status of that individual to generate a binary trait. We choose  $C$  so as to fix the prevalence of the binary trait at 10%. The rate of Type I error and power are estimated at the level of significance 0.05.



In Table X, we consider the scenario where all the traits are distributed as normal and assume that marker locus allele frequency ( $= P(M_1)$ ) is 0.1. Table XI pertains to the scenario where all four traits are distributed as chi-square under the same simulation parameter values considered in Table X. Under this scenario, we find that the power turns out to be more than 99% for most of the simulation parameter values when  $P(M_1) = 0.1$ . Since this phenomenon may be attributed to the similarity between the allele frequencies at the marker locus and QTL, we explore the impact of choosing  $P(M_1) = 0.2$  on power in Table XI. In Table XII, the first associated trait is dichotomized based on  $Y_1$  and converted to a binary trait with the threshold chosen so as to set the prevalence at 10%. The other simulation parameters remain same as in Table X. Similarly, in Table XIII, the first associated trait is binary and is dichotomized based on  $Y_1$ . All other simulation parameters in Table XIII are the same as in Table XI.

In Table XIV, we consider the scenario where the QTL has genetic effects on all four traits, each of which is distributed as chi-square. We assume that the QTL effects on the third and the fourth traits and the genetic heritabilities of these two traits due to the QTL are the same as those for the first and the second traits, respectively. The other simulation parameters remain the same as in Table XI.

### One binary trait

For a binary trait with a recessive mode of inheritance, Wang [2014] showed that *MultiPhen* produces significantly lower power compared to the regular Armitage trend test. We have proved theoretically that *BAMP* is equivalent to the usual allelic case control association test, and hence also asymptotically equivalent to the trend test under the assumption of HWE. Thus, for a sample of 2000 individuals under the assumption of HWE, *BAMP* and the Armitage trend test are expected to be asymptotically equivalent. Thus, we compare the three methods for a binary trait under the same simulation framework considered in Wang [2014] which we briefly describe in the following.

Consider a biallelic locus with two alleles  $A_1$  and  $A_2$  (minor allele). Let  $p_0$ ,  $p_1$ , and  $p_2$  denote the probabilities of the three genotypes  $A_1A_1$ ,  $A_1A_2$ ,  $A_2A_2$ , respectively, and  $f_0$ ,  $f_1$ , and  $f_2$  be their penetrances. Corresponding to this locus, the prevalence of the disease is  $K = p_0f_0 + p_1f_1 + p_2f_2$ . The genotype probabilities in cases are:  $\pi_{0j} = \frac{p_j f_j}{K}$  and those in controls are:  $\pi_{1j} = \frac{p_j(1-f_j)}{(1-K)}$ , for  $j = 0, 1, 2$ . Let  $\gamma_i = \frac{f_i}{f_0}$  be the relative risk of genotype  $i$  to genotype 0. Let  $q$  be  $P(A_2)$ , and so, under HWE,  $p_0 = (1-q)^2$ ,  $p_1 = 2q(1-q)$ ,  $p_2 = q^2$ . A choice of  $(K, q, \gamma_1, \gamma_2)$  fully determine all the parameters needed for data generation, because,

$f_0 = \frac{K}{p_0 + \gamma_1 p_1 + \gamma_2 p_2}$ ,  $f_1 = \gamma_1 f_0$ ,  $f_2 = \gamma_2 f_0$ . Power is calculated based on 10,000 replications and 2000 individuals comprising half of them to be cases and the other half to be controls. We consider the same choices of  $(K, \gamma_1, \gamma_2)$  as considered in Wang [2014] to encompass a variety of modes of inheritance for the binary trait (recessive, dominant, additive). A recessive model is considered by choosing  $\gamma_1 = 1$ , whereas  $\gamma_1 = \gamma_2$  and  $\gamma_1 = (1 + \gamma_2)/2$  represent the dominant and additive models, respectively. The results are provided in Table III. In all the tables, *MPhen* denotes *MultiPhen*.

## Two binary traits

We also consider two binary traits with different modes of inheritance. Different choices of penetrances for the two diseases due to a QTL are considered. A marker locus that is in LD with the QTL is also considered. Different choices of allele frequencies at the loci are made to observe difference in powers among the methods. Different choices of simulation parameters, and type I error & power obtained by three methods for each choice are described in Table IV. All the results are based on 2, 000 individuals and 10, 000 replications.

## Results

### Four traits

For brevity of space, Tables X-XIV are presented in details in the “*Supplementary materials*”. The three methods maintain the desired level of significance (corresponding to  $LD = 0$ ) in all simulation scenarios. With respect to power, we first note some findings observed irrespective of the distribution and type (binary/continuous) of the multiple traits. Of course, the power of detecting association increases as the standardized LD between the marker locus and the QTL increases. We notice that, the power also increases with the degree of correlation among the traits; that is, for a given row in Tables X-XIV with non-zero LD, the power increases with the increase in the value of  $\rho$  (correlation between traits), except for choice 3 in Table XIV. In choice 4, when the associated traits have oppositely directed QTL effects, all the methods result in higher power compared to that for choice 1 (QTL effects in same direction). We also note that, as expected, converting the first associated continuous trait into a binary trait decreases power (Table XII compared to Table X). Similarly, due to the dichotomization of the first continuous trait, the overall power in Table XIII is less in comparison to the powers shown in Table XI.

A summary of the comparative performance by the three methods with respect to the power of detecting multivariate association observed in the Tables X-XIV is provided in Table II. In overall, the allelic tests offer marginal increase of power, in particular, *BAMP* produces a power increase in a range of (1% – 7%) compared to *MultiPhen*.

### One binary trait

From Table III, we observe that, for the recessive model of inheritance of the binary trait, the allelic tests produce much higher power compared to *MultiPhen* with an increase in the range of 10% – 17%. However, for dominant mode of inheritance, *MultiPhen* produces marginally higher power (5% – 6%) compared to the allelic tests. For the additive model, the three methods yield comparable power.

### Two binary traits

We observe from Table IV that, for recessive mode of inheritance of both traits, the allelic tests produce substantially higher power (4% – 19%) over *MultiPhen*. *DAMP* yields a slight increase of power (2% – 5%) compared to *MultiPhen* when both the traits follow an additive mode of inheritance, and one follows additive and the other follows a recessive mode of inheritance. In the later scenario, *BAMP* also produces a marginal increase of power (1% –

3%) in comparison with *MultiPhen*. All three methods produce similar power when the traits follow other combinations of mode of inheritance: both dominant, one dominant and one additive (results not provided for brevity), one recessive and one dominant (results not provided for brevity).

In summary, for four phenotypes, *BAMP* and *DAMP* yield comparable powers, but marginally higher than *MultiPhen*, in particular, *BAMP* yields a consistent marginal increase of power in a range of 1% – 7% compared to *MultiPhen*. For one or two binary traits under a recessive mode of inheritance, the allelic tests offer substantially higher power compared to *MultiPhen* in a range of 4% – 19%.

### Q-Q and test statistics plots

In order to provide a deeper insight of the comparison among the three methods, we present some Q-Q plots and test statistics plots for some selected scenarios from the simulation study.

Some prominent variation in powers are observed in Table XII and XIII related to the situation of mixed (binary and continuous) phenotypes for choice 3 and LD = 0.95. Larger difference in power is observed among the methods for two binary traits with recessive mode of inheritance (Table IV). So, in these three scenarios and for choice 3, LD = 0.66,  $\rho = 0.5$  in Table XI when all four traits are distributed as chi-square, pairwise plots of the test statistics for the three methods are presented in Figure 1 and 2, respectively. Since each of the test statistics follows a chi-square distribution with the same degree of freedom, pairwise plots provide a valid picture of the comparison. For these same choices of simulation scenarios, we consider LD = 0 ( $H_0$ : no association) between the marker locus and QTL, and present Q-Q plots of  $-\log_{10}(\text{pvalue})$  obtained by different methods in Figure 3, 4, 5, 6, provided in the “*Supplementary materials*”.

The Q-Q plots show that, under the null hypothesis of no association, both the allelic tests and *MultiPhen* produce p-values which approximately follow a uniform distribution. From the pairwise plots of the test statistics under the non-null hypothesis of association, we observe that allelic tests, in particular, *BAMP* consistently produces larger values than *MultiPhen*. As expected, the increase for allelic tests is more pronounced in the plot corresponding to two binary traits with recessive mode of inheritance (Figure 2). Thus the Q-Q plots and pair-wise plots of the test statistics in the same simulation scenarios demonstrate the power increase of allelic tests subject to controlling the false positive rate.

### Minimum number of minor alleles to implement *DAMP*

In order for the multivariate CLT to hold in the formulation of *DAMP*, it is necessary that the count of the minor allele at a SNP must be sufficiently large. In Table X, XI, XII, XIII, the minor allele frequency (MAF) at the marker locus is reduced to 0.01 and 0.0125 (for 2000 individuals) in order to induce an average count of the minor allele to be 40 and 50, respectively. For these two choices of low MAF, simulations for choice 1 in Table X, XI, XII, XIII are repeated. For MAF = 0.01, it is found that the type I error by *DAMP* is inflated to 0.06 (at level 0.05) for choice 1 and  $\rho = 0.8$  in Table XI when all the traits follow chi

square distribution. However, in other cases for  $MAF = 0.01$ , the false positive rate by *DAMP* is controlled reasonably well. For  $MAF = 0.0125$ , results for choice 1 in Table XI and XIII are provided in Table V. The results show that the rate of false positive by *DAMP* is overall maintained. So for using *DAMP*, the minor allele count is recommended to be at least 50.

With respect to power, it is observed in Table V that the power by *DAMP* was less than that by the other two methods in a range of 4% – 18%. However, for normally distributed traits and mixed traits comprising one binary and others normal, the three methods yield similar power. We also observe in Table V that, for some cases, the type I error of *MultiPhen* and *BAMP* are slightly inflated. However, for  $MAF = 0.01$  and larger sample size of 5000 individuals, we checked by simulations for choice 1 in Table XI that all the three methods maintain type I error rate appropriately, but power by *DAMP* is smaller than that by *MultiPhen* and *BAMP* in a range of 3% – 10%. Thus, it is more appropriate to implement *DAMP* for common variants rather than variants with lower minor allele frequency.

### Validity of allelic tests

In GWAS, a two-stage procedure in which the markers are first tested for HWE as a data quality-check step and then evaluated for association is performed. In general, a p-value cut-off in the range ( $10^{-3}$  –  $10^{-5}$ ) is used for the genome-wide screening of HWE. Thus, an allelic test can be implemented for a SNP passing through this screening. However, a relevant concern is whether the cut-off of HWE test p-value is sufficient to ensure that HWE holds for a SNP so that the allelic tests can readily be applied. Since it is very difficult to decide on an appropriate cut-off, we explore by simulations the efficiency of the allelic tests in controlling the type I error rate under various degrees of deviation from HWE (in terms of HWE test p-value). We adopt the same simulation design for four traits corresponding to choice 1 and  $\rho = 0.2$  in Table X, XI, and XII, respectively. The results are presented in Table VI. In the last four rows of the table, the QTL has genetic effects on all of the four traits and the simulations corresponding to the first four rows are repeated. For a choice of the genotype frequencies, we generate genotype data at both the marker locus and QTL without assuming HWE. We consider 2000 individuals and zero LD between the marker locus and QTL (no association). Without loss of generality, we also assume the same genotype frequencies at the QTL. The type I error rate is estimated at the level 0.001 based on 20,000 replications. Let the genotype frequencies at the marker locus be denoted by:  $q_0 = P(M_2M_2)$ ,  $q_1 = P(M_1M_2)$ ,  $q_2 = P(M_1M_1)$ , respectively. We consider four different choices of ( $q_0, q_1, q_2$ ) that induces the HWE chi-square test p-value (HWE<sub>pv</sub>) as 0.048, 0.021, 0.0096, and 0.00014, respectively. We observe that both the allelic tests roughly maintain the desired level for the first choice. But for the other three choices, the allelic tests, in particular, *DAMP* seems to produce a significantly inflated rate of false positives when the traits are non-normal. The estimated false positive rates for *MultiPhen* are always less than that obtained for *BAMP* and *DAMP*. Since, the allelic tests maintain the rate of type I error satisfactorily for the first choice, we recommend screening for the HWE-test p-values using the cut-off of 0.05 to implement *BAMP* and *DAMP* at a genome-wide level.

## Combined strategy of *BAMP* and *MultiPhen*

Adopting the commonly used genome-wide significance cut-off  $5 \times 10^{-8}$ , a genome-wide strategy can be to apply *BAMP* for the set of SNPs with a HWE test p-value above 0.05 and to implement a genotype level test (*MultiPhen*) for the SNPs for which the HWE test p-values lie below 0.05 but above the chosen overall HWE screening cut-off (e.g.  $10^{-5}$ ).

For multiple traits, we observed in our simulation study that *BAMP* always produced marginally higher power compared to *MultiPhen*. So for multivariate phenotypes, if we assume that- under HWE- the power of *BAMP* is higher or equal to that of *MultiPhen*, it can be mathematically shown that the power of the combined strategy of implementing *BAMP* under HWE and *MultiPhen* under violated HWE is larger or equal to that by implementing *MultiPhen* alone.

Let  $H_1$  denote that a SNP is associated with the vector of multiple phenotypes. Let  $R_B$  and  $R_M$  denote the events that the SNP association is detected by *BAMP* and *MultiPhen*, respectively. So, the power for *BAMP* =  $P_{H_1}(R_B)$ , and power for *MultiPhen* =  $P_{H_1}(R_M)$ . We have assumed that  $P_{H_1}(R_B|HWE) \geq P_{H_1}(R_M|HWE)$ . Also let  $R_{BM}$  denote that SNP association is detected by the combined strategy of *BAMP* and *MultiPhen*. So, power for the combined strategy =  $P_{H_1}(R_{BM}) = P_{H_1}(R_B|HWE) \times P(HWE) + P_{H_1}(R_M|HWE^c) \times P(HWE^c)$ . Thus for multiple traits, assuming that *BAMP* is as good as *MultiPhen* in terms of power under HWE, the combined strategy of using *BAMP* and *MultiPhen* will produce higher or same power compared to implementing *MultiPhen* alone. Similarly, under  $H_0$  (no association), type I error of the combined strategy =  $P_{H_0}(R_{BM}) = P_{H_0}(R_B|HWE) \times P(HWE) + P_{H_0}(R_M|HWE^c) \times P(HWE^c) = 0.05 \times P(HWE) + 0.05 \times P(HWE^c) = 0.05$ , since *BAMP* maintains the appropriate rate of type I error under HWE.

We note that, in the extreme cases, when HWE holds at all associated SNPs, the power of the combined strategy is the same as that of *BAMP*, and when all the associated SNPs deviate from HWE, the power of the combined strategy becomes the same as *MultiPhen*. For example, in Table X-XIV, we checked by simulations that the powers by combined strategy and *BAMP* are the same since HWE was assumed while simulating the genotype data.

It is clear from the above mathematical expressions that the increase in power by the combined strategy over *MultiPhen* depends on the proportion of associated SNPs at which HWE holds. The expressions reveal that the marginal increase in power of the combined strategy due to *BAMP* will fade away with the increase in proportion of associated SNPs violating HWE. For a Mendelian disorder, the HWE may deviate at most of the causal SNPs due to the effects of selection at the susceptibility loci. However, for a common complex trait due to numerous variants with small effects, HWE may not be violated at a substantial proportion of causal SNPs. In the real data analysis of multiple common complex traits from the GERA cohort given below, the proportion of genome-wide associated SNPs at which HWE holds is substantially high (for this particular data, an empirical estimate is 80%). Thus, due to the possibility that a substantial proportion of SNPs associated with complex traits may not deviate from HWE, the combined strategy of *BAMP* and *MultiPhen* can be

used in a GWAS of multivariate complex traits to take the advantage of consistent marginal increases in power by *BAMP*.

We carry out a simulation study to study the relative performances of the combined strategy and *MultiPhen* without assuming HWE at the associated markers. In order to consider a set of associated markers at some of which HWE can be violated and vice-versa, we select the top 15 GW-significant SNPs identified by *MultiPhen* in the real data analysis of GERA cohort which include 3 SNPs (20%) that yield a significant deviation from the HWE at a level of 0.05. This proportion is consistent with the overall percentage (20%) of GW significant SNPs (using a GW threshold  $5 \times 10^{-8}$ ) deviating from HWE (results reported in Table IX).

The estimated genotype probabilities at the selected 15 SNPs are used to simulate genotypes without assuming HWE in the simulation study. Each of these SNPs is considered as both the QTL and marker, i.e., the marker is in perfect LD with the QTL having the same genotype probabilities as QTL. We choose choice 1 and 3 from Table XI and XIII. All the simulation parameters for generating phenotypes remain the same except the non-genetic variance of every trait being increased by 100 folds in order to enforce the overall (averaged across the 15 SNPs) powers not reaching 100%. The powers that are described in Table VII are estimated at the level 0.05 using 2000 individuals and 2000 replications. For a better understanding, the percentage of replications in which the combined strategy produces a smaller p-value compared to *MultiPhen* is also provided. The results show that the combined strategy produces a marginal gain (1% – 4%) in power compared to *MultiPhen*. It is also observed that, for 58% – 70% of replications across various choices, the combined strategy produces a smaller p-value in comparison with *MultiPhen*. Thus, the simulation study reveals that, with respect to power of detecting multivariate association, there is no loss but a marginal gain by implementing the combined strategy compared to implementing *MultiPhen* alone.

## Application to real data

### Indian population

**Data description**—In a study on hyperhomocysteinemia in a North Indian population, Kumar et al. [2009] analyzed 14 non-synonymous SNPs located in homocysteine metabolism pathway genes. Based on a random sample of 546 individuals selected from the All India Institute of Medical Sciences, two SNPs: *rs9001* in the *CHDH* (Choline Dehydrogenase) gene and *rs1801133* in the *MTHFR* (Methylenetetrahydrofolate Reductase) gene were found to exhibit significant evidence of association (using the Kruskal-Wallis test) with plasma total homocysteine levels, elevated levels of which increase the risk of coronary artery disease (CAD). These associations were validated when an additional 330 individuals from the same population were included in the sample. Moreover, there was a significant interaction of the two SNPs with the nature of diet (vegetarian or non-vegetarian) in modulating the levels of homocysteine. We consider the data on a multivariate phenotype vector comprising three traits: plasma total homocysteine levels, Vitamin B<sub>12</sub> levels (which is significantly negatively correlated with homocysteine levels) and the binary CAD status along with the genotypes at the 14 SNPs for all 876 individuals in the combined sample.

**Analysis and results**—Homocysteine and Vitamin B<sub>12</sub> levels are adjusted for the effects of potential confounders: age, gender, and diet using linear regression. We evaluate the relative performances of the three different methods in detecting association between the genotypes at the 14 SNPs and a multivariate phenotype vector comprising CAD status and the residuals of the other two phenotypes obtained via linear regression. None of the SNPs shows any significant deviation from HWE at a level of significance 0.05. Corresponding to three SNPs: *rs1801133*, *rs12676*, *rs9001*, the p-value produced by at least one of the methods is lower than 0.05. The p-values corresponding to these SNPs for the different methods are presented in Table VIII. For the SNPs *rs1801133* and *rs12676*, both the allelic tests produce significantly lower p-values compared to *MultiPhen*, while for the SNP *rs9001*, the p-values produced by *BAMP* and *MultiPhen* are comparable.

### GERA cohort

**Data description**—We analyzed the “Resource for Genet Epidemiol Research on Adult Health and Aging” (GERA) cohort data obtained from dbGaP [dbGaP Study Accession: phs000674.v1.p1]. We analyzed the European-American individuals which constitutes more than 75% of the full data. The data contains genotype data on 657184 SNPs on 22 autosomal chromosomes for 62318 individuals. The data is composed of 22 case-control phenotypes for the individuals along with adjusting covariates and principal components (PCs) of ancestry. For our analysis, we selected four case-control phenotypes: hypertension (HYPER), diabetes II (DIA2), cardiovascular disease (CARD), and cancers. We also adjusted our analysis for potential covariates: age, gender, smoking status, BMI category, and six PCs.

Following Anderson et al. [2010], we ran filtering of the initial data. First we ran individual-wise screening and removed individuals according to the following criteria: discordant sex information using PLINK [Purcell et al., 2007], missing genotype rate more than 3%, genotype heterozygosity rate beyond the six-sigma limit, and screening for relatedness up to first degree relatives using KING [Manichaikul et al., 2010]. This left us with 59458 individuals. Next, we did per marker screening for: excessive missing genotype rate more than 10%, and screening for SNPs with minor allele frequency (MAF) below 0.01, which left us with 635771 SNPs. In the final analysis, we excluded individuals with missing data for adjusting covariates which left us with 53809 individuals. SNPs were screened for deviation from HWE at a level  $10^{-5}$ .

**Analysis and results**—Since *DAMP* can not adjust for covariates simultaneously in the model, we executed *BAMP* and *MultiPhen* for the joint analysis of the four phenotypes. Adjusting covariates were included along with the phenotypes for both the methods. We also analyzed each of the four binary phenotypes separately by logistic regression of the affection status on genotype adjusting for the same set of covariates incorporated in the joint analysis. Based on the commonly used genome-wide significance level of  $5 \times 10^{-8}$ , 54 SNPs were initially found GW-significant by either *BAMP* or *MultiPhen*. The SNPs were LD filtered for  $r^2 > 0.7$ , which left a total of 30 SNPs associated with at least one of the traits or overall. For these SNPs, results of joint and trait-specific analyses are given in Table IX. The p-values for the chi-square test of HWE are also reported for each of the SNPs.

On chromosome 9, *rs6475606* would not have been picked up by individual phenotype-wise analysis because of each individual trait's p-value fell below the genome-wide significance cut-off. This demonstrates the advantage of employing a joint analysis for multiple phenotypes. We also note that, for 6 SNPs out of the 30 genome-wide significant SNPs, the HWE test p-value falls below 0.05 (underlined), showing the validity of *BAMP* for the majority (80%) of the GW-significant SNPs.

Five SNPs, for which only one of the methods produces a p-value below the genome-wide significance cut-off, are highlighted in Table IX. Among these five SNPs, *rs7896811* and *rs4408545* show significant deviation from HWE at level 0.05, and hence, *BAMP* is not applicable. For the other three SNPs that satisfy HWE, *BAMP* marginally misses the single SNP identified by *MultiPhen* on chromosome 1, and similarly, *MultiPhen* marginally misses the single SNP identified by *BAMP* on chromosome 8. But, *BAMP* produces a genome-wide signal at *rs67279079* on chromosome 11, for which *MultiPhen* yields much weaker signal compared to the genome-wide significance cut-off. Thus, the combined strategy marginally misses the signal on chromosome 1, but identifies both signals on chromosome 8 and 11, which *MultiPhen* fails to pick up. We further note that, for 16 SNPs (66%) among the 24 SNPs at which the HWE holds, *BAMP* produces a smaller p-value compared to *MultiPhen*.

We also map the SNPs in Table IX into genes. Each SNP is located either within the boundaries of a gene or 50KB upstream or downstream of it. Apart from the two SNPs on chromosome 9, every SNP is mapped into one or multiple genes. In Table IX, name of one gene for every SNP is reported. The genes are already known to be associated with a range of phenotypes including blood pressure (hence hypertension), diabetes type II, cardiovascular disease, and multiple type of cancers (breast, lung, prostate, etc.).

## Discussion

We have proposed two allelic tests for association with multiple phenotypes extending the models underlying the methods proposed by O'Reilly et al. [2012] and Lee et al. [2013]. We have investigated their performances compared to an established genotype-level approach *MultiPhen* [O'Reilly et al., 2012] based on simulations and analysis of real data. We also show that *BAMP* is statistically equivalent to the classical case-control test of association for binary traits at the allelic level.

We observe that for four traits, in comparison with *MultiPhen*, *BAMP* yields a consistent marginal increase (1% – 7%) in power for detecting association multiple traits. We also find that, for one or two binary traits under a recessive mode of inheritance, the allelic tests offer much higher power compared to *MultiPhen*. It is also found that *MultiPhen* produces marginally higher power for a single binary trait under dominant mode of inheritance.

While the actual frequencies of causal variants with additive, dominant, and recessive genetic effects are somewhat unknown, additive genetic effects are believed to be the most common and the recessive effects be least common [Henn et al., 2015]. Hence, the



simulations carried out assuming additive genetic effects for the QTL are most reflective of the overall power difference among the methods.

However, we observe that the allelic tests produce a consistent marginal increase in power for four phenotypes and the increase reaches maximum for two binary traits with recessive mode of inheritance. So, for analyzing multivariate phenotypes, there is no risk of loss in power but marginal gain if one employs the combined strategy of *BAMP* and *MultiPhen* instead of implementing *MultiPhen* alone. Our example using the GERA cohort shows some support for this, since *BAMP* picked up the only genome-wide signal on chromosome 11 which *MultiPhen* did not. Both of the real data analyses also demonstrate the possible increase in power in terms of smaller p-values for the majority of top SNPs, that satisfy HWE, produced by the allelic tests in comparison with *MultiPhen*.

For four traits, the simulation results are provided for a single choice of marker allele frequency. However, we performed simulations for other choices of it and observed similar behavior of the methods (results not provided for brevity). *DAMP* offers a substantial computational gain over the other two methods, and hence, is a useful alternative for very large GWAS.

We also observe that, for normally distributed phenotypes, *DAMP* and *BAMP* produce similar power. However, for non-normal traits, *BAMP* yields slightly higher power compared to *DAMP*. It may be due to the fact that, under normality of the phenotypes, the sample means of the multivariate phenotypes within both the groups corresponding to two alleles follow exact multivariate normal distribution, whereas for non-normal traits, the sample means follow multivariate normal asymptotically.

Under the null hypothesis, both the test statistics for *MultiPhen* and *BAMP* are distributed as chi-square with  $k$  degrees of freedom while testing for  $k$  phenotypes. But the number of parameters (intercept) in *BAMP* is one less than that in *MultiPhen* and this is the main reason behind the increase in power.

As opposed to *MultiPhen* and *BAMP*, *DAMP* can not directly incorporate covariates in the model. However, it is a common practice to separately adjust for the relevant covariates before the final genome-wide analysis. Yan et al. [2013] has explored the effects of population structure in the inverted regression model of *MultiPhen* and demonstrated that adjusting for ancestry principal components [Price et al., 2006] in the regression works well to control for population stratification. Since, *BAMP* is the allelic version of *MultiPhen*, it is possible to adjust for population structure in similar way. Considering the above issues and the increased power, for mapping association with multivariate phenotypes, we recommend the hybrid approach of employing *BAMP* and *MultiPhen* depending on the HWE than implementing *MultiPhen* alone.

Apart from the methods considered in this article, some other methods also can incorporate phenotypes of mixed type (discrete and continuous). Liu et al. [2009] proposed a method based on the theory of generalized estimating equations (GEE) primarily for two phenotypes that may comprise a binary phenotype as well. The method can be extended for a larger number of traits. However, a practical limitation of the approach is that the continuous

phenotypes need to be normally distributed which is not desirable for non-normal traits. He et al. [2013] proposed a flexible framework for analyzing multivariate phenotypes. Instead of jointly modeling the phenotypes directly, the joint test statistic is formulated based on the marginal score statistics for the univariate phenotypes accounting for the covariance matrix of the score statistics. They also extended their approach for family based data. They compared their method with TATES [van der Sluis et al., 2013], which provides an overall p-value of association combining p-values for individual traits while adjusting for correlations between them, and found that TATES competes well with their method. We note that Galesloot et al. [2014] compared six different methods of analyzing multivariate phenotypes including *MultiPhen* and TATES. Four methods that were found to perform good include both *MultiPhen* and *TATES*. Since, in our simulation study, the proposed allelic tests perform consistently better than *MultiPhen*, we anticipate that the allelic tests would offer competitive performance in comparison with the methods considered in He et al. [2013] and Galesloot et al. [2014]. We also note that, from both theoretical and computational point of view, the methods considered in this article are simpler compared to those proposed in Liu et al. [2009] and He et al. [2013], and yet inherit the flexibility of incorporating phenotypes with arbitrary distribution.

A notable difference between the models underlying *BAMP* and *MultiPhen* lies in the interpretation of the regression parameters  $(\beta_1, \dots, \beta_k)$  for *BAMP* and  $(\gamma_1, \dots, \gamma_k)$  for *MultiPhen*. The regression parameters in *BAMP* denote the effects of the individual phenotypes on the minor allele frequency at a SNP and hence captures more relevant information compared to the regression parameters in *MultiPhen* which denote the effects of the different phenotypes on the genotype frequencies at the SNP. An additional flexibility of *BAMP* is its statistical basis on a binomial regression model constituted of logistic likelihoods that is much simpler than a proportional odds regression model. It may turn out to be more desirable in a Bayesian framework by placing suitable priors on the regression coefficients since a Bayesian approach involving likelihood based on a logistic regression model rather than a proportional odds regression model is more feasible not only from a methodological but also from a computational point of view.

While *BAMP* and *MultiPhen* would provide significant evidence of association even if a SNP is associated with a subset of phenotypes in the multivariate phenotype vector, an advantage of the methods is that these can be simultaneously used to identify the particular subset. One can individually test for the significance of each phenotype considering the rest of the primary traits as adjusting covariates using a similar likelihood ratio test that is distributed as chi-square with one degree of freedom under the null hypothesis of no association. However, it is important to carry out an appropriate correction for multiple testing (such as Bonferroni, Benjamini-Hochberg, or permutations based procedure) to ensure the appropriate overall false positive rate.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

This work was supported by the Council of Scientific and Industrial Research (CSIR) fellowship 09/093(0112)/2008-EMR-I to Arunabha Majumdar, National Institutes of Health grants R01CA088164, R25CA112355, and the UCSF Goldberg-Benioff Program in Cancer Translational Biology. The GERA cohort data came from a grant, the Resource for Genet Epidemiol 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 Genet Epidemiol 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, 2013. The authors sincerely thank Dr. Shantanu Sengupta for providing access to the data on Coronary Artery Disease and two anonymous reviewers for insightful suggestions that improved the manuscript substantially.

## Appendix

### Proof of property 1

For ease of exposition, we assume that the constituent phenotypes in the multivariate phenotype vector are either all continuous or all binary in nature. HWE is assumed at both the marker and trait loci. We first note that, if  $f_1$ ,  $f_2$ , and  $f_3$  are the joint probability densities (in the case of continuous traits) or joint penetrances (in the case of binary traits) of  $(Y_1, \dots, Y_k)$  conditioned on the three genotypes at the trait locus, it is not possible that  $f_1(y_1, \dots, y_k) = f_2(y_1, \dots, y_k) = f_3(y_1, \dots, y_k)$ , for all possible values of  $(y_1, \dots, y_k)$ . For notational convenience, we denote  $f_1(\cdot) = f_1(y_1, \dots, y_k)$ ,  $f_2(\cdot) = f_2(y_1, \dots, y_k)$ ,  $f_3(\cdot) = f_3(y_1, \dots, y_k)$ . The expressions of genotype probabilities at the marker locus conditioning on the phenotypes are derived in the following.

$$\begin{aligned}
 P(M_1M_1|y_1, \dots, y_k) &= \sum_G P(M_1M_1, \text{ genotype of the individual at the trait locus is } G|y_1, \dots, y_k), \\
 &\text{where the sum is over all possible genotypes at the trait locus:} \\
 G=A_1A_1, A_1A_2, A_2A_2 &= \sum_G P(M_1M_1|G) P(G|y_1, \dots, y_k) \\
 &= \sum_G P(M_1M_1|G) \frac{f(y_1, \dots, y_k|G) P(G)}{\sum_G f(y_1, \dots, y_k|G) P(G)} = \\
 &= \frac{(mp+\delta)^2 f_1(\cdot) + 2(mp+\delta)(m(1-p)-\delta) f_2(\cdot) + (m(1-p)-\delta)^2 f_3(\cdot)}{p^2 f_1(\cdot) + 2p(1-p) f_2(\cdot) + (1-p)^2 f_3(\cdot)}
 \end{aligned}$$

$$\begin{aligned}
 \text{Similarly, } P(M_1M_2|y_1, \dots, y_k) &= \\
 &= \frac{2(mp+\delta)((1-m)p-\delta) f_1(\cdot) + 2[(mp+\delta)((1-m)(1-p)+\delta) + ((1-m)p-\delta)(m(1-p)-\delta)] f_2(\cdot) + 2(m(1-p)-\delta)((1-m)(1-p)+\delta) f_3(\cdot)}{p^2 f_1(\cdot) + 2p(1-p) f_2(\cdot) + (1-p)^2 f_3(\cdot)},
 \end{aligned}$$

$$\begin{aligned}
 \text{and, } P(M_2M_2|y_1, \dots, y_k) &= \\
 &= \frac{((1-m)p-\delta)^2 f_1(\cdot) + 2((1-m)p-\delta)((1-m)(1-p)+\delta) f_2(\cdot) + ((1-m)(1-p)+\delta)^2 f_3(\cdot)}{p^2 f_1(\cdot) + 2p(1-p) f_2(\cdot) + (1-p)^2 f_3(\cdot)}.
 \end{aligned}$$

**If part**

In the logistic link function of *BAMP*.

$$\begin{aligned}\beta_1 &= \dots = \beta_k = 0 \\ \Rightarrow P(M_1|y_1, \dots, y_k) &= \frac{\exp(\alpha)}{1+\exp(\alpha)} = m \\ \Rightarrow \alpha &= \log\left(\frac{m}{1-m}\right) \\ \Rightarrow P(M_1M_1|y_1, \dots, y_k) &= \left\{\frac{\exp(\alpha)}{1+\exp(\alpha)}\right\}^2 = m^2, \\ P(M_1M_2|y_1, \dots, y_k) &= 2m(1-m), \quad \text{and} \quad P(M_2M_2|y_1, \dots, y_k) = (1-m)^2.\end{aligned}$$

Using the previous expressions of the probabilities of the genotypes at the marker locus conditioned on  $(y_1, \dots, y_k)$ , we obtain the following equations:

$$\begin{aligned}\frac{(mp+\delta)^2 f_1(\cdot) + 2(mp+\delta)(m(1-p)-\delta)f_2(\cdot) + (m(1-p)-\delta)^2 f_3(\cdot)}{p^2 f_1(\cdot) + 2p(1-p)f_2(\cdot) + (1-p)^2 f_3(\cdot)} &= m^2 \\ \Rightarrow \delta^2 (f_1(\cdot) - 2f_2(\cdot) + f_3(\cdot)) + 2\delta \{mpf_1(\cdot) + m(1-2p)f_2(\cdot) - m(1-p)f_3(\cdot)\} &= 0, \\ 2m(1-m) &= \frac{2(mp+\delta)((1-m)p-\delta)f_1(\cdot) + 2\{(mp+\delta)((1-m)(1-p)+\delta) + ((1-m)p-\delta)(m(1-p)-\delta)\}f_2(\cdot) + 2(m(1-p)-\delta)((1-m)(1-p)+\delta)f_3(\cdot)}{p^2 f_1(\cdot) + 2p(1-p)f_2(\cdot) + (1-p)^2 f_3(\cdot)} \\ \Rightarrow \delta^2 (f_1 - 2f_2 + f_3) - \delta \{(1-2m)pf_1(\cdot) + (1-2m)(1-2p)f_2(\cdot) - (1-2m)(1-p)f_3(\cdot)\} &= 0, \\ \frac{((1-m)p-\delta)^2 f_1(\cdot) + 2((1-m)p-\delta)((1-m)(1-p)+\delta)f_2(\cdot) + ((1-m)(1-p)+\delta)^2 f_3(\cdot)}{p^2 f_1(\cdot) + 2p(1-p)f_2(\cdot) + (1-p)^2 f_3(\cdot)} &= (1-m)^2 \\ \Rightarrow \delta^2 (f_1(\cdot) - 2f_2(\cdot) + f_3(\cdot)) - 2\delta \{(1-m)pf_1(\cdot) + (1-m)(1-2p)f_2(\cdot) - (1-m)(1-p)f_3(\cdot)\} &= 0.\end{aligned}$$

Combining the last three equations we derive the following equation:

$$\delta \{pf_1(\cdot) + (1-2p)f_2(\cdot) - (1-p)f_3(\cdot)\} = 0, \quad \forall (y_1, \dots, y_k).$$

Now, suppose  $\delta = 0$ . In order to satisfy the above equation,  $pf_1(\cdot) + (1-2p)f_2(\cdot) - (1-p)f_3(\cdot)$  must be zero for all possible choices of  $(y_1, \dots, y_k)$ , and hence all possible values of  $f_1(\cdot)$ ,  $f_2(\cdot)$ ,  $f_3(\cdot)$ . But, there must be some choices of  $(y_1, \dots, y_k)$  for which  $f_1(\cdot)$ ,  $f_2(\cdot)$ ,  $f_3(\cdot)$  are not equal. Hence, to satisfy the condition, it must hold that  $p = 0$ ,  $(1-2p) = 0$ ,  $(1-p) = 0$ , which implies that  $p = 0$ ,  $p = 1/2$ ,  $p = 1$ , simultaneously, and hence a contradiction. The argument is based on the following well known equivalence:  $ax + by + cz = 0, \forall x, y, z \iff a=b=c=0$ .

**Only if part**

Let  $\delta = 0$ . From the expressions of marker locus genotype probabilities conditioning on the values of traits, we obtain:  $P(M_1M_1|y_1, \dots, y_k) = m^2$ ;  $P(M_1M_2|y_1, \dots, y_k) = 2m(1-m)$ ;  $P(M_2M_2|y_1, \dots, y_k) = (1-m)^2$ . It implies that  $P(M_1|y_1, \dots, y_k) = m$ . In *BAMP*,

$$P(M_1|y_1, \dots, y_k) = \frac{\exp\left(\alpha + \sum_{j=1}^k \beta_j y_j\right)}{1 + \exp\left(\alpha + \sum_{j=1}^k \beta_j y_j\right)}. \text{ Thus, } \alpha + \sum_{j=1}^k \beta_j y_j = \log\frac{m}{1-m}, \text{ for all possible } (y_1, \dots, y_k). \text{ Hence, } \beta_j = 0, \forall j = 1, \dots, k.$$

**Proof of property 2**

Suppose  $Y_1$  is a binary trait where  $Y_1$  takes the value 1 for an affected individual and the value 0 for an unaffected individual. Thus, testing for  $\beta_1 = 0$  versus  $\beta_1 \neq 0$  in *BAMP* is

equivalent to testing  $P(M_1|case) = P(M_1/Y_1 = 1) = P(M_1/Y_1 = 0) = P(M_1|control)$  versus  $P(M_1|case) = P(M_1|control)$ .

Next, we derive the expression of the likelihood ratio test (LRT) statistic corresponding to *BAMP*. Suppose, we have  $n_1$  cases and  $n_2$  controls. Let  $n_{1M_1}$  and  $n_{1M_2}$  denote the number of  $M_1$  and  $M_2$  alleles among cases. Similarly, let  $n_{2M_1}$  and  $n_{2M_2}$  denote the number of  $M_1$  and  $M_2$  alleles among controls. Hence,  $n_{1M_1} + n_{1M_2} = 2n_1$  and  $n_{2M_1} + n_{2M_2} = 2n_2$ .

Since the allele frequencies in cases and controls are the same under  $H_0$ , we define

$\pi_0(\alpha) = P(M_1) = \frac{\exp(\alpha)}{1+\exp(\alpha)}$ . Under  $H_1$ , we define the following quantities:

$\pi_1(\alpha, \beta_1) = P(M_1|case) = \frac{\exp(\alpha+\beta_1)}{1+\exp(\alpha+\beta_1)}$  and  $\pi_2(\alpha, \beta_1) = P(M_1|control) = \frac{\exp(\alpha)}{1+\exp(\alpha)}$ . Thus, the LRT statistic is given by:

$$\lambda(\cdot) = \frac{\sup_{\alpha} [\pi_0(\alpha)]^{\binom{n_{1M_1} + n_{2M_1}}{n_{1M_1}}} [1 - \pi_0(\alpha)]^{\binom{n_{1M_2} + n_{2M_2}}{n_{1M_2}}}}{\sup_{\alpha, \beta_1} [\pi_1(\alpha, \beta_1)]^{n_{1M_1}} [1 - \pi_1(\alpha, \beta_1)]^{n_{1M_2}} [\pi_2(\alpha, \beta_1)]^{n_{2M_1}} [1 - \pi_2(\alpha, \beta_1)]^{n_{2M_2}}}$$

Suppose under  $H_0$ ,  $\pi_0(a)$  is denoted as  $p_0$  (hence,  $p_0$  is a 1 – 1 function of  $a$ ). Then, the numerator of  $\lambda(\cdot)$  is the same as:  $\sup_{p_0} p_0^{\binom{n_{1M_1} + n_{2M_1}}{n_{1M_1}}} (1 - p_0)^{\binom{n_{1M_2} + n_{2M_2}}{n_{1M_2}}}$ . Similarly under  $H_1$ , denote  $\pi_1(a, \beta_1)$  and  $\pi_2(a, \beta_1)$ , by  $p_1$  and  $p_2$ , respectively [( $p_1, p_2$ ) is a 1 – 1 transformation of ( $\alpha, \beta_1$ )]. Hence, the denominator of  $\lambda(\cdot)$  is the same as:  $\sup_{p_1, p_2} p_1^{n_{1M_1}} (1 - p_1)^{n_{1M_2}} p_2^{n_{2M_1}} (1 - p_2)^{n_{2M_2}}$  and the LRT statistic simplifies to:

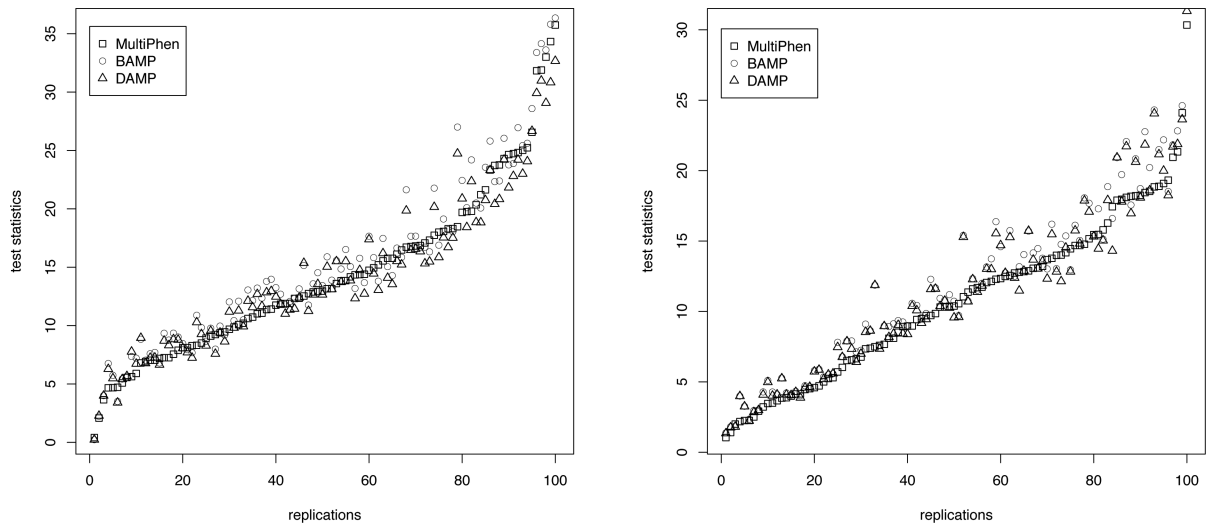
$$\lambda(\cdot) = \frac{\sup_{p_0} p_0^{\binom{n_{1M_1} + n_{2M_1}}{n_{1M_1}}} (1 - p_0)^{\binom{n_{1M_2} + n_{2M_2}}{n_{1M_2}}}}{\sup_{p_1, p_2} p_1^{n_{1M_1}} (1 - p_1)^{n_{1M_2}} p_2^{n_{2M_1}} (1 - p_2)^{n_{2M_2}}}$$

where under  $H_0$ :  $p_0 = P(M_1)$ , and under  $H_1$ :  $p_1 = P(M_1|case)$  and  $p_2 = P(M_1|control)$ . Thus, for a binary trait, testing for association using our model is equivalent to the classical test for difference in allele frequencies among cases and controls.

## References

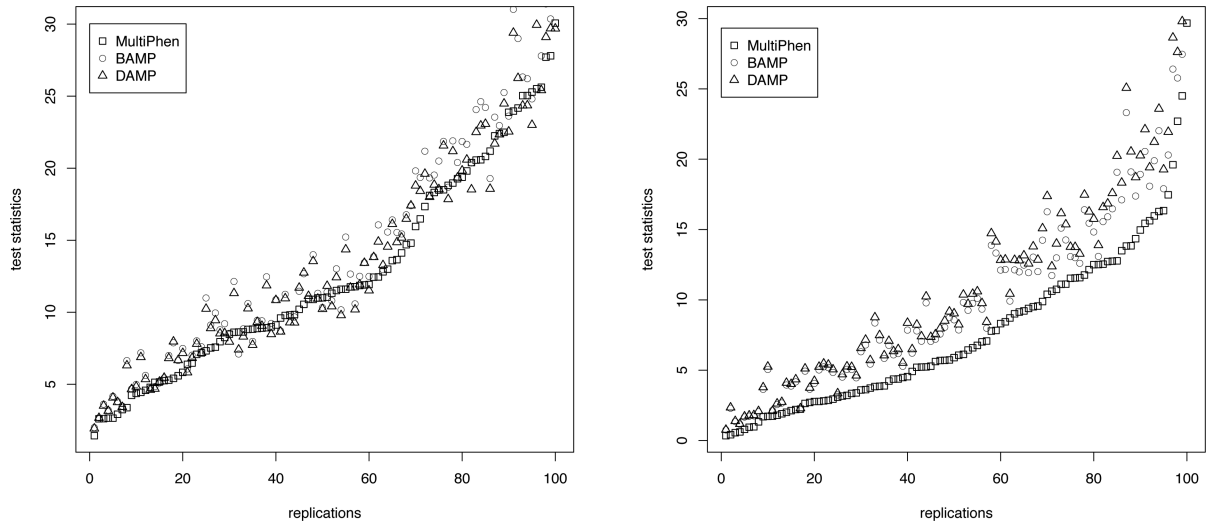
- Almasy L, Blangero J. Variance component methods for analysis of complex phenotypes. *Cold Spring Harb Protoc.* 2010; 2010(5):pdb-top77. [PubMed: 20439422]
- Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc.* 2010; 5(9):1564–1573. [PubMed: 21085122]
- Galesloot TE, Van Steen K, Kiemeny LA, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. *PloS One.* 2014; 9(4):e95923. [PubMed: 24763738]
- Guedj M, Nuel G, Prum B. A note on allelic tests in case-control association studies. *Ann Hum Genet.* 2008; 72(3):407–409. [PubMed: 18355390]
- Guedj M, Wojcik J, Della-Chiesa E, Nuel G, Forner K. A fast, unbiased and exact allelic test for case-control association studies. *Hum Hered.* 2006; 61(4):210–221. [PubMed: 16877868]
- Hartley SW, Monti S, et al. Bayesian methods for multivariate modeling of pleiotropic SNP associations and genetic risk prediction. *Front Genet.* 2012; 3:1–17. [PubMed: 22303408]
- He Q, Avery CL, Lin DY. A general framework for association tests with multivariate traits in large-scale genomics studies. *Genet Epidemiol.* 2013; 37(8):759–767. [PubMed: 24227293]

- Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. *Nature Rev Genet.* 2015
- Klei L, Luca D, Devlin B, Roeder K. Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genet Epidemiol.* 2008; 32:9–19. [PubMed: 17922480]
- Korte A, Vilhjalmsón BJ, et al. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genet.* 2012; 44(9):1066–1071. [PubMed: 22902788]
- Kumar J, Garg G, et al. Single nucleotide polymorphisms in homocysteine metabolism pathway genes association of chdh a119c and mthfr c677t with hyperhomocysteinemia. *Circ Cardiovasc Genet.* 2009; 2(6):599–606. [PubMed: 20031640]
- Lange C, Steen KV, Andrew T, Lyon H, DeMeo DL, Raby B, Murphy A, Silverman EK, MacGregor A, Weiss ST, Laird NM. A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *STAT APPL GENET MOL.* 2004; 3:1544–6115.
- Lee PH, Bergen SE, et al. Modifiers and subtype-specific analyses in whole-genome association studies: a likelihood framework. *Hum Hered.* 2011; 72(1):10–20. [PubMed: 21849790]
- Lee SM, Karrison TG, Cox NJ, Im HK. Quantitative allelic test fast test for very large association studies. *Genet Epidemiol.* 2013; 37(8):831–839. [PubMed: 24185610]
- Liu J, Pei Y, Papasian CJ, Deng HW. Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. *Genet Epidemiol.* 2009; 33:217227.
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 2010; 26(22):2867–2873. [PubMed: 20926424]
- Minica CC, Boomsma DI, Sluis SVD, Dolan CV. Genetic association in multivariate phenotypic data: power in five models. *Twin Res Hum Genet.* 2010; 13(6):525. [PubMed: 21142929]
- Newman WP, Freedman DS, et al. Relation of serum lipoprotein levels and systolic blood pressure to early atherosclerosis. *N Engl J Med.* 1986; 314(3):138–144. [PubMed: 3455748]
- 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]
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association analysis. *Nature Genet.* 2006; 38:904–909. [PubMed: 16862161]
- Purcell S, Neale B, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81(3):559–575. [PubMed: 17701901]
- Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics.* 1997:1253–1261. [PubMed: 9423247]
- Stephens M. A unified framework for association analysis with multiple related phenotypes. *PloS One.* 2013; 8(7):e65245. [PubMed: 23861737]
- Van der Sluis S, Posthuma D, Dolan CV. Tates: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet.* 2013; 9(1):e1003235. [PubMed: 23359524]
- Yan T, Li Q, Li Y, Li Z, Zheng G. Genetic association with multiple traits in the presence of population stratification. *Genet Epidemiol.* 2013; 37(6):571–580. [PubMed: 23740720]
- Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986; 42:121130.
- Zheng G. Can the allelic test be retired from analysis of case-control association studies? *Ann Hum Genet.* 2008; 72(6):848–851. [PubMed: 18652599]
- 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. *Am J Hum Genet.* 2014
- Wang K. Testing Genetic Association by Regressing Genotype over Multiple Phenotypes. *PloS One.* 2014; 9(9):e106918. [PubMed: 25221983]



**Figure 1.**

The figures present the plot of test statistic values obtained by the three methods in 100 replications (ordered in increasing order of the values obtained by *MultiPhen*) for the following two scenarios: in Table XI for  $\rho = 0.5$ , choice 3, LD = 0.66; and in Table XII for  $\rho = 0.8$ , choice 3, LD = 0.95, respectively.



**Figure 2.**

The figures present the plot of test statistic values obtained by the three methods in 100 replications (ordered in increasing order of the values obtained by *MultiPhen*) for the following two scenarios: in Table XIII for  $\rho = 0.8$ , choice 3, LD = 0.95; and in Table IV for two binary traits with recessive mode of inheritance, respectively.



**Table I**

Description of simulation scenarios.  $k$ : total number of traits,  $k_1$ : number of associated traits. In tables X-XIII, since the third and the fourth traits are not associated with the QTL, we impose  $(a_3, b_3) = (a_4, b_4) = (0, 0)$  in each of choices 1-4.

| trait-specific QTL effects and heritabilities |            |                                |                              |                                |                                |
|---|------------|--------------------------------|------------------------------|--------------------------------|--------------------------------|
| Tables  | $(k, k_1)$ | choice 1                       | choice2                      | choice 3                       | choice 4                       |
| X   | (4,2)      | $a_1=1.0, b_1=0.0$             | $a_1=1.0, b_1=0.0$           | $a_1=1.0, b_1=0.0$             | $a_1=1.0, b_1=0.0$             |
|   |            | $a_2=1.0, b_2=0.0$             | $a_2=2.0, b_2=0.0$           | $a_2=1.0, b_2=0.5$             | $a_2=-1.0, b_2=0.0$            |
|   |            | $h_1^2 = 0.5\%, h_2^2 = 0.5\%$ | $h_1^2 = 0.5\%, h_2^2 = 1\%$ | $h_1^2 = 0.5\%, h_2^2 = 0.5\%$ | $h_1^2 = 0.5\%, h_2^2 = 0.5\%$ |
| XI  | (4,2)      | same as Table X                | same as Table X              | same as Table X                | same as Table X                |
| XII   | (4,2)      | In Table X, first              | continuous trait             | is converted to                | a binary trait                 |
| XIII  | (4,2)      | In Table XI, first             | continuous trait             | is converted to                | a binary trait                 |
| XIV   | (4,4)      | $a_1=1.0, b_1=0.0$             | $a_1=1.0, b_1=0.0$           | $a_1=1.0, b_1=0.0$             | $a_1=1.0, b_1=0.0$             |
|   |            | $a_2=1.0, b_2=0.0$             | $a_2=2.0, b_2=0.0$           | $a_2=1.0, b_2=0.5$             | $a_2=-1.0, b_2=0.0$            |
|   |            | $a_3=1.0, b_3=0.0$             | $a_3=1.0, b_3=0.0$           | $a_3=1.0, b_3=0.0$             | $a_3=1.0, b_3=0.0$             |
|   |            | $a_4=1.0, b_4=0.0$             | $a_4=2.0, b_4=0.0$           | $a_4=1.0, b_4=0.5$             | $a_4=-1.0, b_4=0.0$            |
|   |            | $h_1^2 = 0.5\%, h_2^2 = 0.5\%$ | $h_1^2 = 0.5\%, h_2^2 = 1\%$ | $h_1^2 = 0.5\%, h_2^2 = 0.5\%$ | $h_1^2 = 0.5\%, h_2^2 = 0.5\%$ |
|   |            | $h_3^2 = 0.5\%, h_4^2 = 0.5\%$ | $h_3^2 = 0.5\%, h_4^2 = 1\%$ | $h_3^2 = 0.5\%, h_4^2 = 0.5\%$ | $h_3^2 = 0.5\%, h_4^2 = 0.5\%$ |

**Table II**

Summary of the results for the power of detecting multivariate association with four phenotypes by different methods described in Tables X-XIV.

| Tables     | Summary  |
|------------|--|
| Table X    | The traits are normally distributed. Both of the allelic tests consistently produce a marginal 1% – 2% increase in power compared to <i>MultiPhen</i> .  |
| Table XI   | The traits are distributed as chi-square. The performance of <i>MultiPhen</i> is comparable to that of <i>DAMP</i> . However, <i>BAMP</i> performs slightly better than the other two tests offering a 1% – 3% increase of power.  |
| Table XII  | The first associated continuous trait in Table X is converted into a binary trait. In this scenario, <i>MultiPhen</i> and <i>DAMP</i> perform similarly. But, <i>BAMP</i> performs better than <i>MultiPhen</i> producing an increase in power in the range 1% – 5%. For example, in choice 3, for LD=0.95, the power increase is 2% – 5%.   |
| Table XIII | The first associated continuous trait in Table XI is converted into a binary trait. <i>DAMP</i> performs marginally better than <i>MultiPhen</i> while <i>BAMP</i> performs consistently better than the other two methods. It produces higher power than <i>MultiPhen</i> with an increase in the range of 1% – 7%. For example, in choice 3, for LD=0.95, the increase in power of <i>BAMP</i> over <i>MultiPhen</i> is in the range of 5% – 7%. |
| Table XIV  | All the traits are associated with the QTL and distributed as chi-square. While <i>DAMP</i> and <i>MultiPhen</i> produce comparable power, <i>BAMP</i> yields a consistent increase in power compared to <i>MultiPhen</i> , ranging from 1% to 3%.   |

**Table III**

Power comparison for a binary trait.

|   | penetrances         |              | $q = 0.2$   |             |                     | penetrances  |             | $q = 0.3$   |  |
|---|---------------------|--------------|-------------|-------------|---------------------|--------------|-------------|-------------|--|
| $\gamma_1 = 1, \gamma_2 = 1.5$ (recessive)    | $(f_0, f_1, f_2)$   | <i>MPhen</i> | <i>BAMP</i> | <i>DAMP</i> | $(f_0, f_1, f_2)$   | <i>MPhen</i> | <i>BAMP</i> | <i>DAMP</i> |  |
| $K = 0.1$                                     | (0.1,0.1,0.15)      | 0.16         | 0.28        | 0.28        | (0.1,0.1,0.14)      | 0.47         | 0.64        | 0.64        |  |
| $K = 0.01$                                    | (0.01,0.01,0.015)   | 0.14         | 0.24        | 0.24        | (0.01,0.01,0.014)   | 0.40         | 0.54        | 0.54        |  |
| $\gamma_1 = 1.25, \gamma_2 = 1.5$ (additive)  |                     |              |             |             |                     |              |             |             |  |
| $K = 0.1$                                     | (0.09,0.11,0.14)    | 0.87         | 0.88        | 0.88        | (0.09,0.11,0.14)    | 0.93         | 0.93        | 0.93        |  |
| $K = 0.01$                                    | (0.009,0.011,0.014) | 0.80         | 0.80        | 0.80        | (0.009,0.011,0.013) | 0.88         | 0.87        | 0.87        |  |
| $\gamma_1 = 1.25, \gamma_2 = 1.25$ (dominant) |                     |              |             |             |                     |              |             |             |  |
| $K = 0.1$                                     | (0.09,0.11,0.11)    | 0.76         | 0.71        | 0.71        | (0.09,0.11,0.11)    | 0.76         | 0.7         | 0.7         |  |
| $K = 0.01$                                    | (0.009,0.011,0.011) | 0.68         | 0.63        | 0.63        | (0.009,0.011,0.011) | 0.67         | 0.61        | 0.61        |  |

**Table IV**

Power comparison for two binary traits.  $p$  : allele frequency at the QTL,  $m$  : allele frequency at the marker locus. Penetrance vectors  $(f_0, f_1, f_2)$  for the disease 1 and 2 are presented in consecutive rows.

| $p=0.2, m=0.2$ |                    |                  |             |             | $p=0.1, m=0.2$     |                  |             |             |
|----------------|--------------------|------------------|-------------|-------------|--------------------|------------------|-------------|-------------|
| LD             | Penetrances        | <i>MultiPhen</i> | <i>BAMP</i> | <i>DAMP</i> | Penetrances        | <i>MultiPhen</i> | <i>BAMP</i> | <i>DAMP</i> |
| 0 (Terror)     | both recessive     | 0.051            | 0.052       | 0.052       | both additive      | 0.052            | 0.051       | 0.051       |
| 0.33           | (0.25, 0.08, 0.08) | 0.19             | 0.23        | 0.24        | (0.12, 0.08, 0.04) | 0.15             | 0.15        | 0.17        |
| 0.66           | (0.25, 0.08, 0.08) | 0.52             | 0.69        | 0.71        | (0.12, 0.08, 0.04) | 0.47             | 0.47        | 0.50        |
| 0.95           |                    | 0.75             | 0.93        | 0.93        |                    | 0.78             | 0.78        | 0.80        |

| $p=0.1, m=0.3$ |                  |                  |             |             | $p=0.1, m=0.2$      |                  |             |             |
|----------------|------------------|------------------|-------------|-------------|---------------------|------------------|-------------|-------------|
| LD             | Penetrances      | <i>MultiPhen</i> | <i>BAMP</i> | <i>DAMP</i> | Penetrances         | <i>MultiPhen</i> | <i>BAMP</i> | <i>DAMP</i> |
| 0 (Terror)     | both dominant    | 0.052            | 0.051       | 0.051       | recessive, additive | 0.054            | 0.052       | 0.052       |
| 0.33           | (0.1, 0.1, 0.02) | 0.32             | 0.31        | 0.33        | (0.25, 0.08, 0.08)  | 0.11             | 0.11        | 0.12        |
| 0.66           | (0.1, 0.1, 0.02) | 0.88             | 0.87        | 0.88        | (0.12, 0.08, 0.04)  | 0.28             | 0.29        | 0.31        |
| 0.95           |                  | 0.99             | 0.99        | 0.99        |                     | 0.50             | 0.53        | 0.55        |

**Table V**

The minor allele frequency at marker locus is 0.0125. Tterror denotes the type I error and  $\rho$  denotes correlation between traits. Type I error (for LD=0) and power (for non-zero LD) obtained by the three methods are rounded to 3-decimal and 2-decimal points, respectively. Results are based on 2,000 individuals and 10,000 replications.

|                        |       | $\rho = 0.0$ |       |       | $\rho = 0.2$ |       |       | $\rho = 0.5$ |       |       | $\rho = 0.8$ |       |  |
|------------------------|-------|--------------|-------|-------|--------------|-------|-------|--------------|-------|-------|--------------|-------|--|
| LD                     | MPhen | BAMP         | DAMP  | MPhen | BAMP         | DAMP  | MPhen | BAMP         | DAMP  | MPhen | BAMP         | DAMP  |  |
| Choice I in Table XI   |       |              |       |       |              |       |       |              |       |       |              |       |  |
| 0(Tterror)             | 0.05  | 0.05         | 0.054 | 0.053 | 0.054        | 0.052 | 0.052 | 0.052        | 0.052 | 0.051 | 0.051        | 0.054 |  |
| 0.33                   | 0.17  | 0.17         | 0.10  | 0.17  | 0.17         | 0.10  | 0.16  | 0.16         | 0.10  | 0.20  | 0.20         | 0.14  |  |
| 0.66                   | 0.48  | 0.49         | 0.34  | 0.47  | 0.47         | 0.32  | 0.45  | 0.45         | 0.31  | 0.58  | 0.59         | 0.49  |  |
| 0.95                   | 0.77  | 0.78         | 0.64  | 0.77  | 0.77         | 0.64  | 0.75  | 0.75         | 0.62  | 0.88  | 0.88         | 0.83  |  |
| Choice I in Table XIII |       |              |       |       |              |       |       |              |       |       |              |       |  |
| 0(Tterror)             | 0.059 | 0.059        | 0.058 | 0.059 | 0.059        | 0.052 | 0.056 | 0.055        | 0.051 | 0.058 | 0.056        | 0.052 |  |
| 0.33                   | 0.12  | 0.12         | 0.07  | 0.12  | 0.12         | 0.068 | 0.12  | 0.13         | 0.07  | 0.15  | 0.15         | 0.10  |  |
| 0.66                   | 0.30  | 0.29         | 0.17  | 0.29  | 0.29         | 0.16  | 0.29  | 0.29         | 0.17  | 0.42  | 0.42         | 0.32  |  |
| 0.95                   | 0.52  | 0.52         | 0.34  | 0.52  | 0.52         | 0.35  | 0.53  | 0.53         | 0.37  | 0.71  | 0.72         | 0.62  |  |

Type I error rates of the three methods under various degrees of deviation from HWE.  $k_1$ : number of associated traits,  $(q_0, q_1, q_2)$ : genotype frequencies, MAF: minor allele frequency.

**Table VI**

| $k_1 = 2$ | $(q_0, q_1, q_2)$   | MAF   | HWE <sub>pv</sub> | Normal |        |        | Chi-square |        |        | Binary and continuous |        |        |
|-----------|---------------------|-------|-------------------|--------|--------|--------|------------|--------|--------|-----------------------|--------|--------|
|           |                     |       |                   | MPhen  | BAMP   | DAMP   | MPhen      | BAMP   | DAMP   | MPhen                 | BAMP   | DAMP   |
|           | (0.84,0.149,0.011)  | 0.086 | 0.048             | 0.0012 | 0.0014 | 0.0014 | 0.001      | 0.0015 | 0.0016 | 0.0011                | 0.0014 | 0.0015 |
|           | (0.8,0.183,0.017)   | 0.109 | 0.021             | 0.0011 | 0.0016 | 0.0012 | 0.0012     | 0.0016 | 0.0019 | 0.0011                | 0.0017 | 0.0018 |
|           | (0.85,0.139,0.011)  | 0.081 | 0.0096            | 0.0011 | 0.0016 | 0.0009 | 0.0009     | 0.0014 | 0.0021 | 0.0012                | 0.0017 | 0.0019 |
|           | (0.851,0.137,0.012) | 0.08  | 0.00014           | 0.001  | 0.0018 | 0.0013 | 0.0013     | 0.0018 | 0.0021 | 0.0013                | 0.0018 | 0.002  |
| $k_1 = 4$ |                     |       |                   | Normal |        |        | Chi-square |        |        | Binary and continuous |        |        |
|           | (0.84,0.149,0.011)  | 0.086 | 0.048             | 0.0012 | 0.0015 | 0.0014 | 0.0011     | 0.0013 | 0.0016 | 0.0012                | 0.0014 | 0.0015 |
|           | (0.8,0.183,0.017)   | 0.109 | 0.021             | 0.0012 | 0.0015 | 0.0015 | 0.0012     | 0.0015 | 0.002  | 0.0011                | 0.0017 | 0.0019 |
|           | (0.85,0.139,0.011)  | 0.081 | 0.0096            | 0.0011 | 0.0016 | 0.0016 | 0.0008     | 0.0012 | 0.002  | 0.0012                | 0.0017 | 0.002  |
|           | (0.851,0.137,0.012) | 0.08  | 0.00014           | 0.0016 | 0.0016 | 0.0016 | 0.001      | 0.0018 | 0.0025 | 0.0014                | 0.0018 | 0.0019 |

**Table VII**

Comparison of power between *MultiPhen* and combined strategy of *BAMP* and *MultiPhen*. The combined strategy of *BAMP* and *MultiPhen* is denoted by “Comb”, and small.pv presents the percentage of total replications in which the p-value obtained by the combined strategy is smaller than that by *MultiPhen*. Other abbreviations are as follows- all conti: all the four phenotypes are continuous and distributed as chi-square, mixed: one phenotype is binary and other three are continuous and distributed as chi-square.

|           |              | $\rho = 0.0$ |          |              |      | $\rho = 0.2$ |              |      |          | $\rho = 0.5$ |      |          |              | $\rho = 0.8$ |          |              |      |          |  |
|-----------|--------------|--------------|----------|--------------|------|--------------|--------------|------|----------|--------------|------|----------|--------------|--------------|----------|--------------|------|----------|--|
| all conti | <i>MPhen</i> | Comb         | small.pv | <i>MPhen</i> | Comb | small.pv     | <i>MPhen</i> | Comb | small.pv | <i>MPhen</i> | Comb | small.pv | <i>MPhen</i> | Comb         | small.pv | <i>MPhen</i> | Comb | small.pv |  |
| choice 1  | 0.89         | 0.90         | 62%      | 0.88         | 0.89 | 62%          | 0.87         | 0.88 | 62%      | 0.96         | 0.96 | 62%      | 0.96         | 0.96         | 62%      | 0.96         | 0.96 | 62%      |  |
| choice 3  | 0.77         | 0.79         | 66%      | 0.75         | 0.78 | 66%          | 0.76         | 0.78 | 65%      | 0.90         | 0.91 | 66%      | 0.90         | 0.91         | 66%      | 0.90         | 0.91 | 66%      |  |
| mixed     |              |              |          |              |      |              |              |      |          |              |      |          |              |              |          |              |      |          |  |
| choice 1  | 0.65         | 0.67         | 58%      | 0.65         | 0.67 | 58%          | 0.68         | 0.69 | 58%      | 0.88         | 0.89 | 60%      | 0.88         | 0.89         | 60%      | 0.88         | 0.89 | 60%      |  |
| choice 3  | 0.36         | 0.38         | 65%      | 0.36         | 0.38 | 64%          | 0.37         | 0.40 | 65%      | 0.55         | 0.59 | 70%      | 0.55         | 0.59         | 70%      | 0.55         | 0.59 | 70%      |  |

**Table VIII**

P-values for the three SNPs each having a p-value less than 0.05 produced by at least one of the three methods.

| <b>SNP ID</b> | <b><i>rs1801133</i></b> | <b><i>rs12676</i></b> | <b><i>rs9001</i></b> |
|---------------|-------------------------|-----------------------|----------------------|
| <i>MPhen</i>  | 0.072                   | 0.049                 | 0.0047               |
| <i>BAMP</i>   | 0.035                   | 0.018                 | 0.0051               |
| <i>DAMP</i>   | 0.026                   | 0.013                 | 0.0097               |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Table IX

Results for GERA cohort data

| Chrom | rsID        | HWE <sub>pv</sub> | BAMP                   | MultiPhen              | HYPHER                 | DIA2                   | CARD                  | CANCER                 | Gene name |
|-------|-------------|-------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|------------------------|-----------|
| Chr1  | rs 1408420  | 0.51              | 6.20×10 <sup>-8</sup>  | 3.64×10 <sup>-8</sup>  | 0.66                   | 0.30                   | 0.16                  | 3.13×10 <sup>-9</sup>  | RCC2      |
| Chr2  | rs1275988   | 0.81              | 1.38×10 <sup>-9</sup>  | 6.1×10 <sup>-10</sup>  | 1.89×10 <sup>-11</sup> | 0.78                   | 0.84                  | 0.84                   | KCNK3     |
| Chr2  | rs700635    | 0.68              | 4.55×10 <sup>-9</sup>  | 6.21×10 <sup>-9</sup>  | 0.031                  | 0.91                   | 0.67                  | 9.93×10 <sup>-11</sup> | CASP8     |
| Chr2  | rs3769823   | 0.11              | 1.06×10 <sup>-8</sup>  | 4.34×10 <sup>-9</sup>  | 0.17                   | 0.97                   | 0.60                  | 7.54×10 <sup>-11</sup> | ALS2CR12  |
| Chr3  | rs6769511   | 0.16              | 7.65×10 <sup>-9</sup>  | 1.15×10 <sup>-8</sup>  | 0.002                  | 1.30×10 <sup>-9</sup>  | 0.67                  | 0.02                   | MIR548AQ  |
| Chr3  | rs7651090   | 0.06              | 1.04×10 <sup>-8</sup>  | 1.8×10 <sup>-8</sup>   | 0.002                  | 4.19×10 <sup>-9</sup>  | 0.86                  | 0.013                  | IGF2BP2   |
| Chr6  | rs12203592  | <u>0.007</u>      | 3.98×10 <sup>-46</sup> | 1.52×10 <sup>-45</sup> | 0.006                  | 0.4                    | 0.7                   | 1.38×10 <sup>-48</sup> | DUSP22    |
| Chr6  | rs12210050  | 0.51              | 1.0×10 <sup>-14</sup>  | 3.29×10 <sup>-14</sup> | 0.28                   | 0.59                   | 0.39                  | 4.23×10 <sup>-17</sup> | EXOC2     |
| Chr7  | rs1708302   | 0.17              | 4.24×10 <sup>-12</sup> | 6.2×10 <sup>-12</sup>  | 0.11                   | 3.51×10 <sup>-14</sup> | 0.14                  | 0.32                   | JAZF1     |
| Chr7  | rs849138    | 0.22              | 5.13×10 <sup>-11</sup> | 7.15×10 <sup>-11</sup> | 0.051                  | 3.95×10 <sup>-13</sup> | 0.2                   | 0.24                   | JAZF1     |
| Chr7  | rs10274928  | 0.36              | 1.81×10 <sup>-10</sup> | 2.03×10 <sup>-10</sup> | 0.03                   | 2.48×10 <sup>-12</sup> | 0.04                  | 0.32                   | JAZF1     |
| Chr8  | rs 13266634 | 0.97              | 1.10×10 <sup>-8</sup>  | 5.59×10 <sup>-8</sup>  | 0.28                   | 1.30×10 <sup>-9</sup>  | 0.50                  | 0.07                   | SLC30A8   |
| Chr9  | rs6475606   | 0.16              | 2.80×10 <sup>-8</sup>  | 2.70×10 <sup>-8</sup>  | 0.01                   | 0.79                   | 7.35×10 <sup>-8</sup> | 0.004                  |           |
| Chr9  | rs2383207   | 0.19              | 3.16×10 <sup>-8</sup>  | 2.49×10 <sup>-8</sup>  | 0.0002                 | 0.89                   | 2.14×10 <sup>-8</sup> | 0.052                  |           |
| Chr10 | rs4506565   | 0.95              | 2.04×10 <sup>-49</sup> | 4.80×10 <sup>-49</sup> | 0.002                  | 1.51×10 <sup>-55</sup> | 0.17                  | 0.87                   | TCF7L2    |
| Chr10 | rs55899248  | 0.32              | 1.17×10 <sup>-40</sup> | 1.32×10 <sup>-39</sup> | 0.001                  | 5.0×10 <sup>-46</sup>  | 0.99                  | 0.72                   | TCF7L2    |
| Chr10 | rs12255372  | 0.91              | 5.65×10 <sup>-39</sup> | 2.07×10 <sup>-38</sup> | 0.016                  | 8.25×10 <sup>-45</sup> | 0.21                  | 0.83                   | TCF7L2    |
| Chr10 | rs7899529   | 0.73              | 3.93×10 <sup>-19</sup> | 4.36×10 <sup>-19</sup> | 0.008                  | 1.04×10 <sup>-22</sup> | 0.15                  | 0.34                   | TCF7L2    |
| Chr10 | rs61872790  | 0.60              | 1.03×10 <sup>-14</sup> | 6.51×10 <sup>-14</sup> | 0.01                   | 1.08×10 <sup>-17</sup> | 0.85                  | 0.95                   | TCF7L2    |
| Chr10 | rs11196174  | <u>0.00005</u>    | 2.81×10 <sup>-13</sup> | 2.54×10 <sup>-14</sup> | 0.004                  | 1.57×10 <sup>-16</sup> | 0.35                  | 0.79                   | TCF7L2    |
| Chr10 | rs7079711   | 0.77              | 1.44×10 <sup>-10</sup> | 1.24×10 <sup>-10</sup> | 0.86                   | 5.83×10 <sup>-12</sup> | 0.98                  | 0.13                   | TCF7L2    |
| Chr10 | rs12255678  | 0.88              | 1.39×10 <sup>-9</sup>  | 7.58×10 <sup>-10</sup> | 0.003                  | 6.72×10 <sup>-11</sup> | 0.58                  | 0.06                   | TCF7L2    |
| Chr10 | rs 7896811  | <u>0.005</u>      | 1.38×10 <sup>-7</sup>  | 4.35×10 <sup>-8</sup>  | 0.68                   | 6.61×10 <sup>-9</sup>  | 0.9                   | 0.24                   | TCF7L2    |
| Chr11 | rs67279079  | 0.49              | 1.0×10 <sup>-8</sup>   | 2.78×10 <sup>-7</sup>  | 0.69                   | 0.079                  | 0.16                  | 5.15×10 <sup>-10</sup> | TYR       |
| Chr16 | rs12922197  | <u>0.006</u>      | 2.52×10 <sup>-11</sup> | 3.48×10 <sup>-12</sup> | 0.93                   | 0.61                   | 0.76                  | 6.55×10 <sup>-14</sup> | ANKRD11   |
| Chr16 | rs74836424  | 0.52              | 1.4×10 <sup>-10</sup>  | 4.87×10 <sup>-11</sup> | 0.46                   | 0.38                   | 0.73                  | 8.67×10 <sup>-13</sup> | CDK10     |
| Chr16 | rs 4408545  | <u>0.02</u>       | 5.69×10 <sup>-8</sup>  | 4.35×10 <sup>-8</sup>  | 0.30                   | 0.48                   | 0.38                  | 9.90×10 <sup>-10</sup> | CENPBD1   |
| Chr20 | rs4911442   | 0.19              | 9.36×10 <sup>-10</sup> | 1.38×10 <sup>-9</sup>  | 0.33                   | 0.47                   | 0.78                  | 4.69×10 <sup>-12</sup> | NCOA6     |
| Chr20 | rs62212235  | 0.86              | 4.78×10 <sup>-9</sup>  | 4×10 <sup>-9</sup>     | 0.84                   | 0.41                   | 0.38                  | 3.64×10 <sup>-11</sup> | DYNLRB1   |
| Chr20 | rs62211619  | <u>0.0006</u>     | 5.0×10 <sup>-8</sup>   | 2.27×10 <sup>-8</sup>  | 0.90                   | 0.93                   | 0.57                  | 1.26×10 <sup>-10</sup> | ACSS2     |