Association Tests of Multiple Phenotypes: ATeMP

Xiaobo Guo1,2,4, Yixi Li7, Xiaohu Ding6, Mingguang He6, Xueqin Wang1,2,4,5, Heping Zhang1,3,4,*

1 Department of Statistical Science, School of Mathematics & Computational Science, Sun Yat-Sen University, Guangzhou, GD 510275, China, 2 SYSU-CMU Shunde International Joint Research Institute, Shunde, GD 528300, China, 3 Department of Biostatistics, Yale University School of Public Health, New Haven, CT 06520, United States of America, 4 Southern China Research Center of Statistical Science, Sun Yat-Sen University, Guangzhou, GD 510275, China, 5 Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, GD 510080, China, 6 State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, GD 510080, China, 7 Peking University HSBC Business School, Shenzhen, GD 518055, China

* heping.zhang@yale.edu

Abstract

Joint analysis of multiple phenotypes has gained growing attention in genome-wide association studies (GWASs), especially for the analysis of multiple intermediate phenotypes which measure the same underlying complex human disorder. One of the multivariate methods, MultiPhen (O’Reilly et al. 2012), employs the proportional odds model to regress a genotype on multiple phenotypes, hence ignoring the phenotypic distributions. Despite the flexibilities of MultiPhen, the properties and performance of MultiPhen are not well understood, especially when the phenotypic distributions are non-normal. In fact, it is well known in the statistical literature that the estimation is attenuated when the explanatory variables contain measurement errors. In this study, we first established an equivalence relationship between MultiPhen and the generalized Kendall tau association test, shedding light on why MultiPhen can perform well for joint association analysis of multiple phenotypes. Through the equivalence, we show that MultiPhen may lose power when the phenotypes are non-normal. To maintain the power, we propose two solutions (ATeMP-rn and ATeMP-or) to improve MultiPhen, and demonstrate their effectiveness through extensive simulation studies and a real case study from the Guangzhou Twin Eye Study.

Introduction

Genome-wide association studies (GWASs) have emerged as a common tool for identifying the genetic variants for numerous complex diseases. The conventional GWASs focus on a single phenotype, aiming to identify the associations between single nucleotide polymorphisms (SNPs) and a univariate phenotype [1–3]. However, complex human disorders, such as mental disorders, are often characterized by multiple intermediate phenotypes [4, 5]. In addition, many phenotypes, such as body-mass-index and refractive error, are derived from other...
measurements [6, 7]. Modeling the association between multiple phenotypes and a genetic variant may reveal a weak or moderate genetic association that is not apparent from single phenotype GWASs, increasing statistical power and providing fruitful biological insights by identifying pleiotropic variants [8–10].

In recent years we have witnessed an increasing interest in multiple phenotypes GWASs. Among the numerous multivariate methods that have been proposed, some commonly used ones include canonical correlation analysis (CCA) [11], MANOVA [12], and the linear mixed model [13, 14]. However, these methods are highly dependent on the normality assumption, and are known to inflate Type I error [15, 16] when the phenotypic distributions deviate from normality. To deal with this problem, MultiPhen employs the proportional odds model by modeling the genotype score as an ordinal response and the multiple phenotypes as predictors, aiming to identify a combination of phenotypes associated with the genotype. This method ignores the fact that the phenotypes are measured with uncertainty, and hence avoids the need to make a distributional assumption on the phenotypic distributions [16]. Nonetheless, extensive simulations suggest that MultiPhen is one of the most powerful multivariate methods [17].

Despite the promising performance of MultiPhen, the properties of MultiPhen are not well understood. One exception is a recent work by Wang [18] that offered an explicit expression of the score test statistic for MultiPhen and provided some insights into how MultiPhen works in the multiple phenotypes association analyses. Here, we prove that the score test in MultiPhen is in fact equivalent to the generalized Kendall’s tau association test [19], and hence is really an alternative presentation of a method established earlier. Thus, it is not surprising that MultiPhen works well for the multivariate analysis under certain circumstances. Using the equivalence formula to the generalized Kendall’s tau statistic, we demonstrate that MultiPhen may lose power when the phenotypes are non-normal. To maintain robust power, we propose two solutions to improve MultiPhen or the generalized Kendall’s tau when the phenotypes are non-normal.

The rest of this paper is organized as follows. First, we establish the equivalence between MultiPhen and the generalized Kendall’s tau association test, and demonstrate that the MultiPhen may lose power for non-normal phenotypes. Second, we propose two association tests for multiple phenotypes (ATeMP) that perform well even when the phenotypes are non-normal. Finally, extensive simulations and real GWAS data are used to evaluate the performance of ATeMP.

1 Materials and Methods

1.1 Notation

Suppose that there are n subjects in an association study. Let \((Y_i, G_i)\) denote the observed data of the \(i^{th}\) subject, where \(Y_i = (Y_{i1}, \ldots, Y_{iK})^T\) is a vector of K phenotypes of the \(i^{th}\) individual and \(G_i\) is the genotypic score. For simplicity, we consider a single variant and the genotypic score is coded as 0, 1, or 2, corresponding to the number of minor alleles in a biallelic locus.

1.2 MultiPhen

MultiPhen uses the proportional odds logistic regression to model the probability distribution of an individual’s genotype \(G_i\) as a function of the multiple phenotypes,

\[
P(G_i \leq g) = \frac{1}{1 + \exp(-z_0 - \sum_{k=1}^{K} z_k Y_{ik})},
\]  

(1)

Competing Interests: The authors have declared that no competing interests exist.
where the \( \alpha \)'s are regression coefficients. Under this setting, the score test statistic is [18]

\[
S_1 = \frac{1}{n(1 - \pi_0)(1 - \pi_1)(1 - \pi_2)} W^T V^{-1} W,
\]

(2)

where

\[
W = (1 - \pi_0) \sum_{i=0}^n Y_i + (\pi_2 - \pi_0) \sum_{i=2}^n Y_i + (\pi_2 - 1) \sum_{i=2}^n Y_i,
\]

(3)

\[
V = n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})^T,
\]

(4)

and \( \pi_0, \pi_1, \text{and} \pi_2 \) are the proportions of genotype \( G \) with values of 0, 1, and 2, respectively. The statistic \( S \) follows a chi-square distribution with degrees of freedom \( df = K \).

### 1.3 The generalized Kendall’s tau and the equivalence

The generalized Kendall’s tau is one of the earliest association tests for multiple phenotypes [19]. Because it is a nonparametric test, it can be applied to a hybrid of continuous and ordinal phenotypes. Specifically, the generalized Kendall’s tau statistic can be defined as

\[
U = \frac{2}{n(n-1)} \sum_{i<j} f_\xi(G_i - G_j) \begin{pmatrix} f_\xi(Y_i - Y_j) \\ \vdots \\ f_\xi(Y_K - Y_K) \end{pmatrix},
\]

(5)

where \( f_\xi(\cdot) \) and \( f_\eta(\cdot) \) are kernel functions. Two popular choices of the kernel function are the identity function and the sign function. For clarity, let \( f_\xi \) be the identity function because \( G \) is in an ordinal scale, and let \( f_\eta \) be the sign function. Then, statistic \( U \) can be simplified as

\[
U = \frac{2}{n(n-1)} \sum_{i<j} \text{sign}(G_i - G_j)(Y_i - Y_j) \propto \sum_{i=1}^n \bar{g}_i Y_i,
\]

(6)

where

\[
\bar{g}_i = \frac{1}{n} \sum_{j=1}^n \text{sign}(G_i, G_j) = \begin{cases} 1 - \pi_0 & \text{if } G_i = 0, \\ \pi_2 - \pi_0 & \text{if } G_i = 1, \\ \pi_2 - 1 & \text{if } G_i = 2. \\ \end{cases}
\]

(7)

Conditional on the phenotypes, the generalized Kendall’s tau test statistic can be constructed as [19]

\[
S_2 = U^T \text{var}(U|Y) U = (\sum_{i=1}^n \bar{g}_i Y_i)^T \text{var}(\sum_{i=1}^n \bar{g}_i Y_i) (\sum_{i=1}^n \bar{g}_i Y_i).
\]

(8)

Note that \( \sum_{i=1}^n \bar{g}_i Y_i = W \) defined in Eq (3), and as shown in the appendix,

\[
\text{var}(\sum_{i=1}^n \bar{g}_i Y_i) = (1 - \pi_0)(1 - \pi_1)(1 - \pi_2) \sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})^T,
\]

(9)

therefore the generalized Kendall’s tau test statistic \( S_2 \) is equal to the score test statistic \( S_1 \) of MultiPhen. Given the earlier work on the generalized Kendall’s tau, it is not surprising that
MultiPhen works well for the multiple phenotypes association studies under various circumstances.

1.4 ATeMP

The MultiPhen used the classic technique in genetic analysis [20] by conditioning on the phenotypes, and avoided the need to assume phenotypic distributions. However, when the phenotypes are non-normal, MultiPhen may lose power. This is more convenient to see by examining the generalized Kendall’s tau. For example, when all phenotypes are continuous, the identity function is the most natural choice for the kernel function. It is known that this choice is not efficient in the absence of normality [21]. To maintain the power for testing the non-normally distributed phenotypes, we introduce two solutions for association tests of multiple phenotypes (ATeMP):

- ATeMP-rn: The idea is to replace the original phenotypes with their normalized ranks, a common approach to transforming non-normal data [14, 22]. Let \((R_{1k}, \ldots, R_{nk})\) be the rank vector of the \(k\) dimensional phenotypic vector \((Y_{1k}, \ldots, Y_{nk})\). Next, we can employ the inverse normal transformation, and transform \(Y_{ik}\) into
  \[
  \frac{Y_{ik}}{\Phi^{-1}\left(\frac{R_{ik}}{n+1}\right)}.
  \]
  Then, we apply the MultiPhen or equivalently generalized Kendall’s tau.

- ATeMP-or: For a non-normally distributed phenotype, we employ the ordinal residual transformation as described above, and transform \(Y_{ik}\) into
  \[
  \frac{1}{n} \sum_{j=1}^{n} \text{sign}(Y_{jk} - Y_{ik}).
  \]
  Then, we apply the MultiPhen or equivalently generalized Kendall’s tau.

1.5 Simulation Study 1: Bivariate Phenotypes

We conducted simulation studies to systematically evaluate the efficiency as well as the robustness of ATeMP. We generated bivariate traits under the bivariate linear model

\[
Y_{i1} = \beta_{G1} * G_i + \beta_{E1} * E_i, \quad (12)
\]

\[
Y_{i2} = \beta_{G2} * G_i + \beta_{E2} * E_i + \epsilon_i, \quad (13)
\]
where \( G_i \) is the causal variant with minor allele frequency of 0.2, \( E_i \) is a random effect, and \( \epsilon \) is the random error following \( N(0, \sigma^2) \). Varying the distribution of \( E_i \) among several non-normal distributions yields a variety of non-normal phenotypes. Specifically, we set \( \beta_{G1} = 0.1 \) and \( \beta_{G2} = 0, \) or 0.05, or 0.1, and considered the following different distributions for \( E_i \): (1) \( N(0, 1) \), (2) \( t(3) \), (3) Laplace(1.5, 1) and (4) Gamma(1, 2). We chose suitable values of \( \beta_{E1} \), \( \beta_{E2} \) and \( \sigma^2 \) such that the variances of both \( Y_{i1} \) and \( Y_{i2} \) are equal to 1 and the between-phenotype correlation, \( r \), varies from -0.8 to 0.8 in an increment of 0.4.

To evaluate the statistical power, we simulated 1000 datasets under each simulation scenario above. Each simulated dataset consisted of 2000 unrelated individuals. The significance level was fixed at \( 5 \times 10^{-4} \). This nominal level of significance is much higher than the typical level of significance in GWAS to reduce the computational time in simulation. However, we believe it is small enough for the purpose of comparing the power of MultiPhen, ATeMP-rn, and ATeMP-or.

We assessed the Type I error of these tests by letting MAF be 5%. 50000 datasets were simulated and the significance level was set to be \( 5 \times 10^{-4} \) in this simulation study. To assess the asymptotic approximation, we also considered relatively small sample sizes of 300 and 500.

1.6 Simulation Study 2: High Dimensional Phenotypes

To further evaluate the efficiency and robustness of ATeMP, we considered high dimensional phenotypes. The phenotypes are generated using a linear additive model

\[
Y_k = \beta_k G + \sqrt{a} U_k + \sqrt{1 - a} \epsilon_k, \quad k = 1, \cdots, K, \tag{14}
\]

where \((U_1, \cdots, U_K)^T\) follows multivariate normal distribution with mean 0 and covariance matrix \( \Sigma \). A gradient of strong to low levels of correlation for \( \Sigma \) is simulated; that is, \( \rho_{ij} = 0.8^{i-j} \). Under the alternative hypothesis, we assumed that the genetic variant is associated with one third of the phenotypes. We simulated independent \( \epsilon_k \) from one of the following distributions: (1) \( N(0, 1) \); (2) \( t(3) \); (3) Laplace(1.5, 1); (4) Gamma(1, 2). Finally, \( a \) was set to be 0.4 and the number of phenotypes \( K \) was set to be 5 and 10.

To evaluate the statistical power, we simulated 1000 datasets under each simulation scenario above. Each simulated dataset consisted of 1000 unrelated individuals. The significance level was fixed at \( 5 \times 10^{-4} \). The minor allele frequency of the causal variant \( G \) is set to be 0.3. The genetic variant explains 0.3% of the phenotypic variations when \( \epsilon_k \) follows the normal distribution, and 0.6% for the other distributions. We assessed the Type I error by simulating 50000 datasets, and the sample sizes were set to be 300, 500 and 1000.

1.7 Study of Myopia: Testing Candidate SNPs from Guangzhou Twin Project

Here, we applied MultiPhen, ATeMP-rn, and ATeMP-or to evaluated 38 candidate SNPs which are identified from three large GWASs [3, 24, 25] for refractive error. We analyzed a dataset from the Guangzhou Twin Eye Study, which iss a population-based registry designed to examine the genetic and environmental etiologies for myopia. It was launched in 2006, and has completed eight consecutive annual follow-up examinations, with more than 1200 twin pairs participating. In brief, twins born in Guangzhou aged 7 to 15 years received annual eye examinations from 2006 and on. The protocol and examination procedures have been published elsewhere [26]. Written, informed consent was obtained for all participants from either parents or guardians of the participating children after careful explanation of the study in detail, including the discussion and specific consent for the use of DNA information. Ethical committee approval was obtained from the Zhongshan University Ethical Review Board and
Ethics Committee of Zhongshan Ophthalmic Center [26]. We focus on refractive error, which is the most common eye disorder in the world and is the leading cause of blindness [3]. Spherical lens (SPH) and cylindrical lens (CYL), two major intermediate traits of refractive error, have gained increasing interest in the GWAS [27]. Borrowing the strength of the multiple phenotypes association studies, in this report, we are interested in the multiple phenotypes associations analysis for SPH and CYL. Fig 1 displays the distributions of SPH and CYL. We can observe that the distribution of CYL is heavily skewed, suggesting that transformed phenotypes would be preferable before performing the association tests. Specifically, we employed both the inverse normal transformation and the ordinal residual transformation for CYL and SPH.

The current data are from the Guangzhou Twin Eye Study. A detailed description has been published elsewhere [26]. The GWAS data included 1055 individuals from the first-born twins. Age and gender were considered as covariates.

2 Results

2.1 Simulation Studies of Statistical Power and Type I Error

Fig 2 presents the power comparison under different simulation settings for bivariate phenotypes. We can learn from Fig 2 that MultiPhen can lose a great deal of power when the phenotypes are non-normal. The loss is more severe, as shown in Fig 2, when the phenotypes are heavily skewed such as from the Gamma distribution. However, ATeMP-rn and ATeMP-or can recover the loss. Table 1 displays the results of power comparisons under different simulation settings when the number of phenotypes are five and ten. Similarly to the power comparison for bivariate phenotypes, ATeMP-rn and ATeMP-or can recover the power loss when the phenotypes are non-normal. These simulations confirm that transforming non-normal phenotypes is necessary. Even though MultiPhen makes no assumption on the phenotypic distributions, it does not necessarily mean that it is efficient.

To offer a practical guide, we summarize the order of superiority between different methods. When the phenotypic distribution is heavily-tailed, such as the t distribution or the Laplace distribution, ATeMP-or is the most powerful approach in all of the considered simulation settings as can be seen clearly from Fig 2 and Table 1. When the phenotypic distribution is heavily skewed, such as the Gamma distributions, ATeMP-rn is the preferred method for the bivariate phenotypes. However, the performance of ATeMP-rn and ATeMP-or is almost the same when the phenotypes are high dimensional, such as five or ten in our simulation studies.

Table 2 reports the Type I error rates when the nominal significance level is set to be $5 \times 10^{-4}$. We can observe that the Type I error rates of ATeMP-rn and ATeMP-or are very
close to the nominal values, indicating that these methods can control Type I error well in the considered simulation settings. The Type I error rates of MultiPhen is inflated for the \( t \) distribution when the sample size is 300 or 500. We do not observe inflated Type I error rate for MultiPhen when the sample size is 2000. S1 Table also presents the Type 1 error rate when the number of phenotypes are 5 and 10. We can observe that all methods can control Type I error well in the considered simulation settings, indicating that the asymptotic distribution provides an adequate approximation for high dimensional phenotypes.

**Table 1.** The power of the multiple phenotypes association tests at the significance level \(5 \times 10^{-4}\) when the number of phenotypes are 5 and 10.

| No. Phenotypes | Distribution | MultiPhen | ATeMP-nr | ATeMP-or |
|----------------|--------------|-----------|-----------|-----------|
| 5              | normal       | 0.63      | 0.63      | 0.55      |
|                | \( t \)      | 0.18      | 0.24      | 0.28      |
|                | Laplace      | 0.25      | 0.29      | 0.31      |
|                | Gamma(1,2)   | 0.23      | 0.36      | 0.37      |
| 10             | normal       | 0.73      | 0.72      | 0.63      |
|                | \( t \)      | 0.56      | 0.75      | 0.80      |
|                | Laplace      | 0.41      | 0.45      | 0.48      |
|                | Gamma(1,2)   | 0.38      | 0.52      | 0.52      |

doi:10.1371/journal.pone.0140348.t001
2.2 Association Study on Myopia

In Table 3, we display the SNPs with p-value < 0.05 from the joint analysis. ATeMP-rn yields nearly the same results as ATeMP-or, and the most significant SNP (rs12229663 with p-value of $4.9 \times 10^{-4}$) is identified by the ATeMP-or. For the SNPs with p-value < 0.01, most of the p-values from ATeMP are smaller than those from MultiPhen, suggesting again that transforming phenotypes is helpful in this real data analysis. These results confirm the observations from the simulation studies. For SNPs with p-value > 0.01 (Table 3 and S2 Table), there are no apparent benefits from ATeMP.

After the Bonferroni correction, no SNPs are significant by using MultiPhen. However, ATeMP-rn or ATeMP-or identified one significant SNP rs12229663.

### 3 Discussion

In this report, we first pointed out and prove that a recent method for multiple phenotypes association testing, MultiPhen, is in fact equivalent to an earlier test proposed for the same purpose. After establishing this equivalence, we demonstrated that MultiPhen suffers from a substantial loss of power when the phenotypic distributions were non-normal. This calls for the caution that the use of a distribution-free test may be convenient, but it may also be inefficient.

To recover the power loss of MultiPhen, we proposed two phenotypic transformations prior to the use of MultiPhen or the equivalent generalized Kendall’s tau. The first method, ATeMP-rn, employs the frequently used inverse normal transformation for the non-normal phenotypes.
before any association test. The second method, ATeMP-or, uses a particular form of residuals in a proportional odds model involving an ordinal response [23, 28]. Extensive simulations demonstrate that ATeMP tests can recover the power when the phenotypic distributions are heavy-tailed or highly-skewed, while MultiPhen suffers from a substantial loss of power. In addition, we also compared the power by using the permutation method rather than the asymptotic distribution. The results (S1 Fig) indicate again that transforming phenotypes is helpful when the phenotypic distributions are non-normal.

In our simulation studies, we observed that the power of the multivariate methods is high when the correlation of bivariate phenotypes is negative and the genetic effects on the individual phenotypes are positive. Others [13, 16, 29] have also noted this phenomenon that the power increases when the correlation of the phenotypes is in opposite direction to the phenotypic genetic effects. It can also be explained from the perspective of principle component analysis [29].

We applied MultiPhen and ATeMP tests to evaluate 38 candidate SNPs from the Guangzhou Twin Eye Study. Five SNPs showed nominally significant p-value (p-value < 0.05), indicating that part of candidate SNPs of refractive error are associated with its two major intermediate traits. Our real data analysis confirmed that ATeMP tests are superior to MultiPhen, underscoring the usefulness of transforming the non-normal phenotypes prior to association testing, despite the fact that MultiPhen is distribution-free.

Appendix: The derivation of $\text{var}(\sum_{i=1}^{n} \tilde{g}_i Y_i)$

We first note that

$$\sum_{i=1}^{n} \tilde{g}_i = n((1 - \pi_{\theta})\pi_0 + (\pi_2 - \pi_0)(1 - \pi_0 - \pi_2)) + (\pi_2 - 1)\pi_2) = 0$$

and

$$\text{var}(\tilde{g}_i) = (1 - \pi_{\theta})^2\pi_0 + (\pi_2 - \pi_0)^2(1 - \pi_0 - \pi_2) + (\pi_2 - 1)^2\pi_2$$

$$= (1 - \pi_{\theta})(1 - \pi_1)(1 - \pi_2)$$

since $\pi_0 + \pi_1 + \pi_2 = 1$. Therefore,

$$\text{var}(\sum_{i=1}^{n} \tilde{g}_i Y_i) = \text{var}(\sum_{i=1}^{n} \tilde{g}_i (Y_i - \bar{Y}))$$

$$= \sum_{i=1}^{n} (Y_i - \bar{Y})\text{var}(\tilde{g}_i)(Y_i - \bar{Y})^T$$

$$= (1 - \pi_{\theta})(1 - \pi_1)(1 - \pi_2)\sum_{i=1}^{n} (Y_i - \bar{Y})(Y_i - \bar{Y})^T.$$

Supporting Information

S1 Fig. The power of the multiple phenotypes association tests at the significance level $5 \times 10^{-4}$ under different simulation settings. Different types of curve represent different methods. The simulation settings are the same as the simulation studies for bivariate phenotypes in Section 1.5. To alleviate the computational burden, the sample size was set to be 500, and the significance level was set to be 0.05. (EPS)
S1 Table. Type I error of the multiple phenotypes association tests when the phenotypes are five and ten, respectively. The nominal significance level is set to be $5 \times 10^{-4}$, and the sample sizes are set to be 300, 500 and 100, respectively.

(XLS)

S2 Table. P-values from association tests of 38 candidate SNPs by jointly analyzing CYL and SPH.

(XLS)

Acknowledgments

Zhang’s research is partially supported by the U.S. National Institute on Drug Abuse (R01 DA016750), a 1000-plan scholarship from the Chinese Government, and the International Collaborative Research Fund from NSFC(11328103). Guo’s research is supported by the NSFC (11401600), and the Fundamental Research Funds for the Central Universities (15lgpy07). Wang’s research is partially supported by the free application projects from the SYSU-CMU Shunde International Joint Research Institute, NSFC(11271383) and Chinese Government and the International Collaborative Research Fund from NSFC(11328103).

Author Contributions

Conceived and designed the experiments: XG XW HZ. Performed the experiments: XG. Analyzed the data: XG XW HZ. Contributed reagents/materials/analysis tools: YL XD MH. Wrote the paper: XG YL XD MH XW HZ.

References

1. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nature genetics. 2007; 39(7):870–874. doi:10.1038/ng2075 PMID: 17529973
2. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. science. 2007; 316 (5829):1341–1345. doi:10.1126/science.1142382 PMID: 17463248
3. Verhoeven VJ, Hysi PG, Wojciechowski R, Fan Q, Guggenheim JA, Höhn R, et al. Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. Nature genetics. 2013; 45(3):314–318. doi: 10.1038/ng.2554 PMID: 23396134
4. Guo X, Liu Z, Wang X, Zhang H. Genetic association test for multiple traits at gene level. Genetic epidemiology. 2013; 37(1):122–129. doi: 10.1002/gepi.21688 PMID: 23032486
5. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460(7256):748–752. doi:10.1038/nature08185 PMID: 19571811
6. Lavery J, Gibson J, Shaw D, Rosenthal A. Refraction and refractive errors in an elderly population. Ophthalnic and Physiological Optics. 1988; 8(4):394–396. PMID:3258361
7. Speilotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature genetics. 2010; 42 (11):937–948. doi:10.1038/ng.686 PMID: 20935630
8. Amos C, Laing A. A comparison of univariate and multivariate tests for genetic linkage. Genetic epidemiology. 1993; 10(6):671–676. doi: 10.1002/gepi.1370100657 PMID: 8314079
9. Schifano ED, Li L, Christiani DC, Lin X. Genome-wide association analysis for multiple continuous secondary phenotypes. The American Journal of Human Genetics. 2013; 92(5):744–759. doi: 10.1016/j.ajhg.2013.04.004 PMID: 23643383
10. Zhu W, Zhang H. Why do we test multiple traits in genetic association studies? Journal of the Korean Statistical Society. 2009; 38(1):1–10. doi: 10.1016/j.jkss.2008.10.006 PMID: 19665045
11. Ferreira MA, Purcell SM. A multivariate test of association. Bioinformatics. 2009; 25(1):132–133. doi: 10.1093/bioinformatics/btn563 PMID: 19019849
12. Suo C, Toulopoulou T, Bramon E, Walshe M, Picchioni M, Murray R, et al. Analysis of multiple phenotypes in genome-wide genetic mapping studies. BMC bioinformatics. 2013; 14(1):151. doi: 10.1186/1471-2105-14-151 PMID: 23639181
13. Korte A, Vilhjálmsson BJ, Segura V, Plat A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nature genetics. 2012; 44(9):1066–1071. doi: 10.1038/ng.2376 PMID: 22902788
14. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature methods. 2014; 11(4):407–409. doi: 10.1038/nmeth.2848 PMID: 24531419
15. Medland SE, Neale MC. An integrated phenomic approach to multivariate allelic association. European Journal of Human Genetics. 2010; 18(2):233–239. doi: 10.1038/ejhg.2009.133 PMID: 19707246
16. O’Reilly PF, Hoggart CJ, Pomyen Y, Calboli FC, Elliott P, Jarvelin MR, et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS One. 2012; 7(5):e34861. doi: 10.1371/journal.pone.0034861
17. Galesloot TE, Van Steen K, Kiemeney LA, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. PloS one. 2014; 9(4):e95923. doi: 10.1371/journal.pone.0095923 PMID: 24763738
18. Wang K. Testing Genetic Association by Regressing Genotype over Multiple Phenotypes. PloS one. 2014; 9(9):e106918. doi: 10.1371/journal.pone.0106918 PMID: 25221983
19. Zhang H, Liu CT, Wang X. An association test for multiple traits based on the generalized Kendall’s tau. Journal of the American Statistical Association. 2010; 105(490):473–481. doi: 10.1198/jasa.2009.ap08387 PMID: 20711441
20. Haseman J, Elston R. The investigation of linkage between a quantitative trait and a marker locus. Behav Genet. 1972; 2:3–19. doi: 10.1007/BF01066731 PMID: 4157472
21. Hollander M, Wolfe DA, Chicken E. Nonparametric statistical methods. John Wiley & Sons; 2013.
22. Wei C, Li M, He Z, Vasevolozhskaya O, Schaid DJ, Lu Q. A Weighted U-Statistic for Genetic Association Analyses of Sequencing Data. Genetic epidemiology. 2014; 38(8):689–708. doi: 10.1002/gepi.21864 PMID: 25331574
23. Li C, Shepherd BE. A new residual for ordinal outcomes. Biometrika. 2012;p. asr073.
24. Kiefer AK, Tung JY, Do CB, Hinds DA, Mountain JL, Francke U, et al. Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. PLoS Genet. 2013; 9(2):e1003299. doi: 10.1371/journal.pgen.1003299 PMID: 23468642
25. Cheng CY, Schache M, Ikram MK, Young TL, Guggenheim JA, Vitart V, et al. Nine loci for ocular axial length identified through genome-wide association studies, including shared loci with refractive error. The American Journal of Human Genetics. 2013; 93(2):264–277. doi: 10.1016/j.ajhg.2013.06.016 PMID: 24144296
26. Zheng Y, Ding X, Chen Y, He M. The Guangzhou Twin Project: An Update. Twin Research and Human Genetics. 2013; 16(01):73–78. doi: 10.1017/thg.2012.120 PMID: 23186635
27. Li Q, Wojciechowski R, Simpson CL, Hysi PG, Verhoeven VJ, Ikram MK, et al. Genome-wide association study for refractive astigmatism reveals genetic co-determination with spherical equivalent refractive error: the CREAM consortium. Human genetics. 2015; 134(2):131–146. doi: 10.1007/s00439-014-1500-y PMID: 25367360
28. Zhang H, Wang X, Ye Y. Detection of Genes for Ordinal Traits in Nuclear Families and a Unified Approach for Association Studies. Genetics. 2006; 172:693–699. doi: 10.1534/genetics.105.049122 PMID: 16219774
29. Aschard H, Vilhjálmsson BJ, Greлицhe N, Morange PE, Trégouët DA, Kraft P. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. The American Journal of Human Genetics. 2014; 94(5):662–676. doi: 10.1016/j.ajhg.2014.03.016 PMID: 24746957
Author/s:
Guo, X; Li, Y; Ding, X; He, M; Wang, X; Zhang, H

Title:
Association Tests of Multiple Phenotypes: ATeMP

Date:
2015-10-19

Citation:
Guo, X., Li, Y., Ding, X., He, M., Wang, X. & Zhang, H. (2015). Association Tests of Multiple Phenotypes: ATeMP. PLOS ONE, 10 (10), https://doi.org/10.1371/journal.pone.0140348.

Persistent Link:
http://hdl.handle.net/11343/257053

File Description:
published version

License:
CC BY