



# A modified generalized Fisher method for combining probabilities from dependent tests

Hongying Dai<sup>1,2,3\*</sup>, J. Steven Leeder<sup>2,4</sup> and Yuehua Cui<sup>5</sup>

<sup>1</sup> Department of Pediatrics, Research Development and Clinical Investigation, Children's Mercy Hospital, Kansas City, MO, USA

<sup>2</sup> Department of Pediatrics, University of Missouri-Kansas City, Kansas City, MO, USA

<sup>3</sup> Department of Informatic Medicine and Personalized Health, University of Missouri-Kansas City, Kansas City, MO, USA

<sup>4</sup> Department of Pediatrics, Clinical Pharmacology and Therapeutic Innovation, Children's Mercy Hospital, Kansas City, MO, USA

<sup>5</sup> Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA

## Edited by:

José M. Álvarez-Castro,  
Universidade de Santiago de  
Compostela, Spain

## Reviewed by:

Wei Hou, Stony Brook University,  
USA

J. Concepcion Loredó-Osti,  
Memorial University, Canada

## \*Correspondence:

Hongying Dai, Department of  
Pediatrics, Research Development  
and Clinical Investigation, Children's  
Mercy Hospital, 2401 Gillham Road,  
Kansas City, MO 64108, USA  
e-mail: hdai@cmh.edu

Rapid developments in molecular technology have yielded a large amount of high throughput genetic data to understand the mechanism for complex traits. The increase of genetic variants requires hundreds and thousands of statistical tests to be performed simultaneously in analysis, which poses a challenge to control the overall Type I error rate. Combining  $p$ -values from multiple hypothesis testing has shown promise for aggregating effects in high-dimensional genetic data analysis. Several  $p$ -value combining methods have been developed and applied to genetic data; see Dai et al. (2012b) for a comprehensive review. However, there is a lack of investigations conducted for dependent genetic data, especially for weighted  $p$ -value combining methods. Single nucleotide polymorphisms (SNPs) are often correlated due to linkage disequilibrium (LD). Other genetic data, including variants from next generation sequencing, gene expression levels measured by microarray, protein and DNA methylation data, etc. also contain complex correlation structures. Ignoring correlation structures among genetic variants may lead to severe inflation of Type I error rates for omnibus testing of  $p$ -values. In this work, we propose modifications to the Lancaster procedure by taking the correlation structure among  $p$ -values into account. The weight function in the Lancaster procedure allows meaningful biological information to be incorporated into the statistical analysis, which can increase the power of the statistical testing and/or remove the bias in the process. Extensive empirical assessments demonstrate that the modified Lancaster procedure largely reduces the Type I error rates due to correlation among  $p$ -values, and retains considerable power to detect signals among  $p$ -values. We applied our method to reassess published renal transplant data, and identified a novel association between B cell pathways and allograft tolerance.

**Keywords:** generalized Fisher method (Lancaster procedure), weight function, correlated  $p$ -values, multiple hypothesis testing, high dimensional genetic data

## INTRODUCTION

Rapid developments in molecular technology have created high throughput data in search of genetic variants associated with complex traits. As the cost of experiments goes down, the amount of data that can be generated, and the resulting complexity of statistical analysis required to interpret the data goes up. The increase of genetic variants requires more statistical testing to be performed simultaneously, which poses a challenge to control the genome wide Type I error rate. False discovery rate (FDR) and its extended methods have been proposed to adjust  $p$ -values in multiple tests in order to control the genome wide Type I error (Benjamini and Hochberg, 1995; Cheng and Pounds, 2007). However, in large-scale hypothesis testing, these methods often require very a large sample size to maintain power of detecting risk factors.

The global test (also named omnibus test) of  $p$ -values can combine evidence and turn dimensionality from a curse into rich information. From a systems biology perspective, genes, cells,

tissues, and organs function as a system through metabolic networks and cell signaling networks. In non-Mendelian inheritance patterns, such as complex disorders, a subset of genetic variants may jointly confer moderate effects in mediating molecular activities. As a result, signals may not be significant in single marker-single trait analysis, but many such values from related genes might provide valuable information on gene function and regulation. For instance, in pathway analysis (Khatri et al., 2012) and gene set enrichment analysis (Subramanian et al., 2005), multiple genes that work together to serve a particular biological function are often analyzed jointly as a gene set. Several pathway repositories, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2004), PANTHER classification system for protein sequence data (Nikolsky and Bryant, 2009), and Reactome pathways in humans (Matthews et al., 2009) have been established, and are continually being updated. For non-Mendelian diseases and complex traits, identification of isolated genetic variants is insufficient to summarize the complex

association with disease. The “most-significant SNPs/genes” approach often detects variants with small effect sizes and odds ratios ranging between 1.3 and 2 (Wacholder et al., 2004). Therefore, integrating information from pathways, gene sets, and networks will provide useful information in understanding the gene regulation mechanism. Furthermore, filtration techniques can be integrated with global testing of  $p$ -values to remove sets of genetic variants that are not related to traits, and thereby reduce the dimensionality of the data (Dai and Charnigo, 2008; Dai et al., 2012a).

The global test of  $p$ -values evaluates the pattern (distribution) of  $p$ -values instead of selecting  $p$ -values less than an arbitrary threshold. Therefore, this method has the potential to identify multiple genes with small effects. If we assume that all individual tests are independent and arise from genetic variants with no effects, then  $p$ -values are identically and independently distributed as Uniform (0, 1). Taking this as a null hypothesis for the pattern of  $p$ -values in the global test, one can assess whether  $p$ -values, especially small  $p$ -values, are generated by chance. The global test of  $p$ -values is robust and can be applied to  $p$ -values from varying statistical models including  $t$ -tests, analysis of variance (ANOVA), linear mixed models, and so forth. Multiple simulation studies and case studies have demonstrated that this approach usually has sufficient power to detect signals of genetic association from a group of genes. For instance, Peng et al. (2010) has assessed Fisher’s combination test and Sidak’s combination test, Sime’s combination test and the FDR method using 13 published genome wide association studies (GWAS), and the results indicate that combined  $p$ -value approaches can identify biologically meaningful pathways associated with the disease susceptibility. A review of methods of global test of  $p$ -values, developmental trends and their application to genetic data analysis has been presented by (Dai et al., 2012b).

One category of global tests of  $p$ -values involves combining  $p$ -values in the form of  $\sum_i H(p_i)$ , where  $p$ -values might first be transformed by a function  $H$ . So far, several statistical methods have been developed to combine  $p$ -values. Let  $p_i (i = 1, 2, \dots, n)$  be independent  $p$ -values obtained from  $n$  hypothesis tests. Under the null hypothesis ( $H_0$ ) that  $p$ -values follow a Uniform (0, 1) distribution, Fisher (1932) shows that  $-2 \sum_{i=1}^n \ln(p_i)$  follows a chi-square distribution with  $2n$  degrees of freedom. For a one sided test with a nominal error rate of  $\alpha$ , one can reject the null hypothesis when the test statistics exceeds the  $(1 - \alpha) * 100\%$  percentile of  $\chi_{2n}^2$ . Stouffer (Stouffer et al., 1949) proposed a  $z$ -test by transforming  $p$ -values to standard normal variables, i.e.,  $\sum_{i=1}^n \frac{\Phi^{-1}(1-p_i)}{\sqrt{n}}$ , where  $\Phi^{-1}$  is the inverse Cumulative Distribution Function (CDF) for  $N(0, 1)$ . Under the null hypothesis, the  $z$ -test statistic follows  $N(0, 1)$ .

Although there is no consensus regarding the most powerful method of combining  $p$ -values, Littell and Folks (1971, 1973) demonstrated that the Fisher’s method of combining independent tests is asymptotically Bahadur efficient (Bahadur, 1967). Subsequently, weighting schemes have been incorporated into the Fisher’s method and the  $z$ -test. Lancaster (1961) generalized the Fisher method by converting independent  $p$ -values to chi-square variables with  $w_i$  degrees of freedom and he

showed that  $\sum_{i=1}^m \gamma_{(w_i/2, 2)}^{-1} (1 - p_i) \sim \chi_d^2$ ,  $d = \sum_i w_i$  under  $H_0$ , where  $\gamma_{(w_i/2, 2)}^{-1}$  is the inverse CDF of Gamma distribution. Mosteller and Bush (1954) proposed a weighted  $z$ -test,  $\sum_i w_i \Phi^{-1} (1 - p_i) / \sqrt{\sum_i w_i^2}$ , which follows  $N(0, 1)$  under  $H_0$ .

In a separate paper, we have proved that the Lancaster procedure achieves the optimal Bahadur efficiency. We further demonstrated that the Lancaster procedure yields higher Bahadur efficiency than the weighted  $z$ -test. The Bahadur efficiency ratio gives the limiting ratio of sample sizes required by two statistics to attain an equally small significance level. Thus, Bahadur efficiency is an important method to compare test statistics. From the perspective of Bahadur efficiency, the Lancaster procedure asymptotically requires a relatively smaller sample size than other weighted  $p$ -value combining methods. This prompted us to focus on modification of the Lancaster procedure for correlated genetic data in this work.

Although the Fisher’s method and Lancaster procedure both achieve the optimal Bahadur efficiency, the Lancaster procedure is more general and can be viewed as a generalized Fisher’s method with weighting functions. There are three advantages to carefully select appropriate weight functions in genetic data analysis. Firstly, weight functions allow incorporation of prior biological information. Genetic data are complex and can be measured from different sources. Thus, weight functions can be used as a tool to incorporate meaningful information from different sources in order to interpret and derive biological insight from gene expression profiles. (Wu and Lin, 2009) provides a review of statistical methods for analysis of microarray data by incorporating prior biological knowledge using gene sets and biological pathways, which consist of groups of biologically similar genes. They show that the use of prior knowledge has led to a better understanding of the biological mechanisms underlying phenotypic responses. Secondly, weight functions can be used to remove bias. For instance, larger genes may contain more probes and/or SNPs. Therefore, larger genes will exert a stronger influence on the  $p$ -value combining methods as compared to smaller genes (Wang et al., 2007). To avoid this bias, one can consider a weight function to adjust for gene size when combining  $p$ -values. We will illustrate this approach in sections Empirical Assessments and Case Study: Renal Transplant Tolerance Data. Thirdly, as suggested by Benjamini and Hochberg (1997), Genovese et al. (2006), procedures that assign weights positively associated with the underlying alternative hypotheses will usually improve power. Therefore, one needs to carefully choose an appropriate weight function, either based on the biological knowledge, or by statistical hypotheses. An arbitrary weight is inappropriate for the Lancaster procedure.

In this work, we will provide modifications to the Lancaster procedure to accommodate correlation structures among  $p$ -values. The proposed method provides a generalization to the Fisher’s method with a weight function and can be used in pathway analysis and gene sets enrichment analysis for a variety of genetic data including microarray gene expression data, GWAS data, and next generation sequencing data. In essence, investigators first dissect genetic variants by biological functions

or prior knowledge, then combine the  $p$ -values from these gene sets to identify whether a proportion of genetic variants are associated with traits.

### CORRELATED LANCASTER PROCEDURES

In this section, we allow  $p$ -values to be correlated. Consider a Lancaster test statistic  $T = \sum_{i=1}^n \gamma_{(w_i/2,2)}^{-1} (1 - p_i)$  where  $\gamma_{(w_i/2,2)}^{-1}$  is the inverse CDF of Gamma distribution with a shape parameter  $w_i/2$  and a scale parameter 2. This transformation converts  $p_i \sim \text{Uniform}(0, 1)$  to a chi-square distribution, i.e.,  $\gamma_{(w_i/2,2)}^{-1} (1 - p_i) \sim \chi_{w_i}^2$  where  $\chi_{w_i}^2$  is a chi-square distribution with  $w_i > 0$  degree(s) of freedom. The parameter  $w_i$  serves as a weight function to adjust the individual  $p$ -values. When  $p$ -values are independent,  $T$  has an exact chi-square distribution with  $\sum_{i=1}^n w_i$  degrees of freedom.

For correlated  $p$ -values,  $T = \sum_{i=1}^n \gamma_{(w_i/2,2)}^{-1} (1 - p_i)$  does not follow  $\chi_{\sum_{i=1}^n w_i}^2$ . The distribution of  $T$  does not have an explicit analytical form. To address this issue, we consider a Satterthwaite approximation by approximating a scaled  $T$  statistic with a new chi-square distribution (Li et al., 2011). Let  $cT \approx \chi_{\nu}^2$  where  $c > 0$  is a scalar and  $\nu > 0$  is the degree of freedom for the approximated chi-square distribution. Note that

$$\begin{aligned}
 E(T) &= E\left(\sum_{i=1}^n \gamma_{(w_i/2,2)}^{-1} (1 - p_i)\right) = \sum_{i=1}^n w_i \text{ and} \\
 \text{Var}(T) &= \text{var}\left(\sum_{i=1}^n \gamma_{(w_i/2,2)}^{-1} (1 - p_i)\right) \\
 &= \sum_{i=1}^n \text{var}\left(\gamma_{(w_i/2,2)}^{-1} (1 - p_i)\right) \\
 &\quad + 2 \sum_{i < j} \text{cov}\left(\gamma_{(w_i/2,2)}^{-1} (1 - p_i), \gamma_{(w_j/2,2)}^{-1} (1 - p_j)\right) \\
 &= 2 \sum_{i=1}^n w_i + 2 \sum_{i < j} \rho_{ij},
 \end{aligned}$$

where  $\rho_{ij} = \text{cov}\left(\gamma_{(w_i/2,2)}^{-1} (1 - p_i), \gamma_{(w_j/2,2)}^{-1} (1 - p_j)\right)$  takes the correlations among  $p$ -values into account.

We propose the following five approaches to approximate the distribution of  $T$ . In approximation (A), we use the Satterthwaite method to match the mean and variance of  $cT$  and  $\chi_{\nu}^2$ , and then solve the equations to derive  $c$  and  $\nu$ . Koziol (1996) have proposed multiple methods to approximate the Lancaster procedure, but these approximations require the assumption of independence. In approximation (B)–(E), we extend the work of Koziol (1996) to correlated data by first approximating  $cT$  with  $\chi_{\nu}^2$  then approximating  $\chi_{\nu}^2$  using varying methods.

- $T_A$  approximation. Correlation among  $p$ -values is taken into consideration, and then Satterthwaite's approximation is used (Patnaik, 1949) to derive new degrees of freedom:

$$T_A = cT \approx \chi_{\nu}^2, \text{ where } c = \frac{\nu}{E(T)} \text{ and } \nu = 2 \frac{[E(T)]^2}{\text{var}(T)}.$$

- $T_B$  approximation.  $cT$  is first approximated by  $\chi_{\nu}^2$ , followed by Fisher's approximation (Fisher, 1922) to  $\chi_{\nu}^2$ :

$$T_B = \sqrt{2 \frac{\nu T}{E(T)}} \approx N(\sqrt{2\nu - 1}, 1).$$

- $T_c$  approximation. After approximating  $cT$  by  $\chi_{\nu}^2$ , the Wilson–Hilferty approximation is performed (Wilson and Hilferty, 1931) to derive  $\chi_{\nu}^2$ .

$$\text{Let } T_c = \sqrt[3]{\frac{T}{E(T)}}, \text{ then } T_c \approx N\left(1 - 2/(9\nu), \sqrt{2/(9\nu)}\right).$$

- $T_D$  approximation. Approximate  $cT$  by  $\chi_{\nu}^2$ , followed by the Cornish–Fisher expansion (Fisher and Cornish, 1960) to  $\chi_{\nu}^2$ . Let  $x_{\alpha}$  denote the  $\alpha$ -percentage point of the standard normal distribution, that is,  $\Phi(x_{\alpha}) = \alpha$ . It follows that the corresponding percentage point for  $T_D = \frac{\nu T}{E(T)}$  is given by

$$\begin{aligned}
 &\nu + \sqrt{2\nu}x_{\alpha} + \frac{2}{3}(x_{\alpha}^2 - 1) + \frac{x_{\alpha}^3 - 7x_{\alpha}}{9\sqrt{2\nu}} - \frac{6x_{\alpha}^4 + 14x_{\alpha}^2 - 32}{405\nu} \\
 &\quad + \frac{9x_{\alpha}^5 + 256x_{\alpha}^2 - 433x_{\alpha}}{4860\nu\sqrt{2\nu}}.
 \end{aligned}$$

- $T_E$  approximation. Approximate  $cT$  by  $\chi_{\nu}^2$  then perform saddle point approximation (Lugannani and Rice, 1980) to  $\chi_{\nu}^2$ . Let  $T_E = \frac{T}{E(T)}$ . Then  $\Pr(Y_E \leq y) = \Phi(a_y) - \phi(b_y^{-1} - a_y^{-1})$  for  $y \neq 1$  and  $\Pr(Y_E \leq 1) = 0.5 - (3\sqrt{\pi\nu})^{-1}$ , where  $a_y = \sqrt{2\nu(yt_y - K(t_y))\text{sign}(t_y)}$ ,  $b_y = t_y\sqrt{\nu K''(t_x)}$  and  $K(t) = -0.5 \log(1 - 2t)$ , and  $t_y = (y - 1)/2y$ .

When the covariance  $\rho_{ij}$  is unknown, one can use the permutation approach to estimate  $\rho_{ij}$  by shuffling the phenotype variable among subjects. For the  $k$ th permutation ( $k = 1, 2, \dots, m$ ), we keep the genetic variants within the subject to preserve the correlation structure, then randomly assign the phenotype variable to subjects. Individual hypothesis testing can be done on all  $n$  genetic variants separately to generate the  $p$ -value vector  $p^k = (p_1^k, p_2^k, \dots, p_n^k)^t$ . The permutation is repeated  $m = 1000$  times, and  $\rho_{ij}$  is estimated from  $(p^1, p^2, \dots, p^m)$ .

The accuracy of the five approximate distributions to the correlated Lancaster procedure is then assessed using  $p$ -values with varying correlation structures. We consider six different types of correlation structures, including fixed and random compound symmetric as well as random positive definite variance-covariance structures for  $\Sigma$ . Let  $I$  be an identity matrix,  $\mathbf{1}$  be a vector of 1 s,  $\otimes$  be the Kronecker product, and superscript  $t$

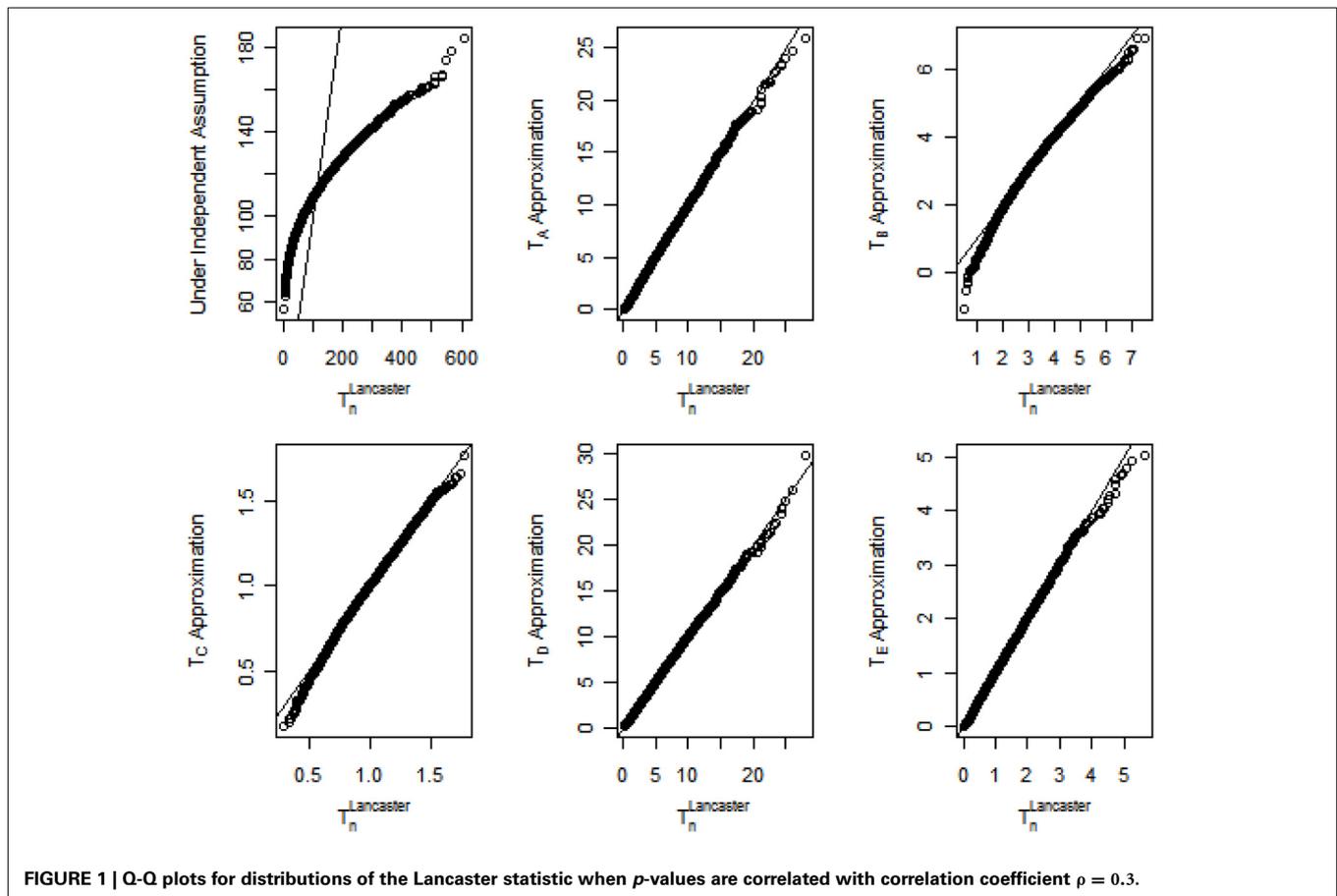
be the transposition. In Cases I–V, let  $\Sigma = \text{Block} \otimes I_{20}$  be compound symmetric variance matrices with 20 blocks of size 5 where  $\text{Block} = \bar{1}_5 \bar{1}_5^T \rho + (1 - \rho)I_5$ . We vary  $\rho$  over two fixed values with  $\rho = 0.3$  for moderate dependence and  $\rho = 0.6$  for strong dependence. In addition, we simulate random correlation coefficients from beta and uniform distributions, i.e.,  $\rho \sim \beta(0.3, 1.5)$  and  $\rho \sim \text{uniform}(-0.2, 0.2)$ , which ensures that 20 variance blocks have distinct correlation coefficients  $\rho$  within  $\Sigma$ . More generally, we consider random positive definite correlation matrices  $\Sigma$  that vary across samples and simulation runs.

The quantile-quantile (Q-Q) plot assessing the accuracy of the proposed methods when the correlation coefficient  $\rho = 0.3$  is shown in **Figure 1**. For clarity, the Lancaster statistic  $T$  that combines  $n$   $p$ -values is renamed as  $T_n^{\text{Lancaster}}$  in **Figure 1**. For the original Lancaster procedure under the independence assumption, the general trend of the Q-Q plot is flatter than the reference line  $y = x$ , indicating the limiting distribution for the test statistic in the original Lancaster procedure is less dispersed than the distribution of  $T_n^{\text{Lancaster}}$  under correlation structures. As a result, the original Lancaster procedure will have severely inflated Type I errors. In contrast, the five approximations ( $T_A, \dots, T_E$ ) match the underlying distribution of  $T_n^{\text{Lancaster}}$ . For data with stronger internal correlation,  $T_A, T_D$ , and  $T_E$  better approximate  $T_n^{\text{Lancaster}}$ . The Q-Q plots under other correlation structures are similar to

**Figure 1**. To save space, these similar results are not shown, but can be provided upon request.

## EMPIRICAL ASSESSMENTS

We assess the Type I error rates and power for the proposed correlated Lancaster procedures and compare them to the independent Lancaster procedure (Lancaster, 1961). SNPs from a pathway of haploid GWAS are simulated using linkage disequilibrium (LD) (Li et al., 2011). Let  $q_1$  and  $q_2$  be the minor allele frequencies (MAFs) at loci 1 and 2. Assuming Hardy–Weinberg equilibrium, the genotype at locus 1 can be randomly generated using a binomial distribution. Given the distribution of SNP at locus 1, one can simulate the genotype at locus 2. To do so, let  $D$  be a measure of LD. Then the conditional probability for the genotype at locus 2 given the genotype at locus 1 can be expressed as  $P(A|B) = [q_A q_B + D]/q_B$ ,  $P(a|B) = [(1 - q_A)q_B - D]/q_B$ ,  $P(A|b) = [q_A(1 - q_B) - D]/(1 - q_B)$ , and  $P(a|b) = [(1 - q_A)(1 - q_B) + D]/(1 - q_B)$  where  $A$  and  $B$  represent the minor alleles at the two loci. For a diploid genome, similar idea can be applied and the simulation details can be found at Cui et al. (2008). We simulate a pathway with 5 genes with varying numbers of SNPs in each gene listed in parenthesis i.e., G1(12), G2(8), G3(5), G4(3), G5(2). The MAF of each SNP was set to be 0.3. We simulate different levels of LD for SNPs from



the same gene with  $D = 0, 1.5, 2$ , and uniform(0, maximum of LD). The variable  $D = 0, 1.5$ , and  $2$  suggests no LD, moderate LD, and very strong LD among SNPs with the corresponding correlation  $R = 0, 0.71$ , and  $0.95$ . Six scenarios for disease susceptibility ( $p$ ) are simulated

- Case I:  $\ln(p/(1-p)) = \beta_1 G_{1,2} + \beta_2 G_{1,5} + \beta_3 G_{1,7} + \beta_4 G_{1,8} + \beta_5 G_{1,12}$ .
- Case II:  $\ln(p/(1-p)) = \beta_1 G_{2,2} + \beta_2 G_{2,4} + \beta_3 G_{2,6} + \beta_4 G_{3,2} + \beta_5 G_{3,3}$ .
- Case III:  $\ln(p/(1-p)) = \beta_1 G_{3,2} + \beta_2 G_{3,4} + \beta_3 G_{4,1} + \beta_4 G_{4,3} + \beta_5 G_{5,1}$ .
- Case IV:  $\ln(p/(1-p)) = \beta_1 G_{1,1} + \beta_2 G_{1,3} + \beta_3 G_{1,7} + \beta_4 G_{1,8} + \beta_5 G_{1,10} + \beta_6 G_{1,11} + \beta_7 G_{1,12}$ .
- Case V:  $\ln(p/(1-p)) = \beta_1 G_{3,1} + \beta_2 G_{3,3} + \beta_3 G_{4,2} + \beta_4 G_{3,2} + \beta_5 G_{4,3} + \beta_6 G_{5,1}$ .
- Case VI:  $\ln(p/(1-p)) = \beta_1 G_{1,2} + \beta_2 G_{2,2} + \beta_3 G_{3,3} + \beta_4 G_{5,2} + \beta_5 G_{1,5} + \beta_6 G_{1,7} + \beta_7 G_{3,3} + \beta_8 G_{5,1}$ .

**Table 1 | Type I error and power for independent Lancaster Procedure and five approximations to correlated Lancaster Procedures when sample size = 200 and linkage disequilibrium  $D = 0.15$ .**

	Independent Lancaster procedure	$T_A$	$T_B$	$T_C$	$T_D$	$T_E$
<b>CASE I</b>						
$\beta = 0$	<i>0.101</i>	0.038	0.042	0.039	0.039	0.038
$\beta = 0.4$	0.999	0.995	0.995	0.995	0.995	0.995
$\beta = 0.6$	1	1.000	1	1	1	1
<b>CASE II</b>						
$\beta = 0$	<i>0.1</i>	0.037	0.041	0.038	0.038	0.037
$\beta = 0.4$	0.947	0.863	0.875	0.864	0.865	0.863
$\beta = 0.6$	0.997	0.995	0.995	0.995	0.995	0.995
<b>CASE III</b>						
$\beta = 0$	<i>0.078</i>	0.038	0.038	0.038	0.038	0.038
$\beta = 0.4$	0.735	0.506	0.522	0.508	0.507	0.506
$\beta = 0.6$	0.961	0.864	0.876	0.866	0.866	0.863
<b>CASE IV</b>						
$\beta = 0$	<i>0.107</i>	0.046	0.051	0.046	0.047	0.046
$\beta = 0.4$	0.997	0.997	0.997	0.997	0.997	0.997
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE V</b>						
$\beta = 0$	<i>0.084</i>	0.036	0.038	0.037	0.037	0.036
$\beta = 0.4$	0.884	0.71	0.724	0.71	0.711	0.71
$\beta = 0.6$	0.989	0.952	0.957	0.953	0.953	0.952
<b>CASE VI</b>						
$\beta = 0$	<i>0.084</i>	0.036	0.038	0.037	0.037	0.036
$\beta = 0.4$	0.741	0.57	0.585	0.572	0.572	0.568
$\beta = 0.6$	0.953	0.898	0.904	0.898	0.898	0.898

A weight function is applied to adjust for the gene size\*.  
 \*The nominal error rate is set to be 0.05. Type I error rates are listed when  $\beta = 0$ . Power is listed when  $\beta > 0$ . Inflated Type I error rates are italicized.  
 \*A weight function  $w_i = 2/\sqrt{n_i}$  is applied to each test to adjust for the size of gene.

Weight functions can be used to remove potential bias when combining  $p$ -values. Wang et al. (2007) and others have noted that larger genes contain more probes and/or SNPs. Therefore, larger genes may exert a stronger influence on the  $p$ -value combining methods compared to smaller genes. To avoid this bias, we set the weight function  $w_i = 2/\sqrt{n_i}$  where  $n_i$  is the number of SNPs in the  $i$ th gene. When  $n_i = 1$ ,  $Y_{(w_i/2, 2)}^{-1}(1-p_i)$  transforms  $p$ -value into a variable with  $\chi_2^2$  distribution.

We simulate data with sample sizes  $n = 200$  (Tables 1, 4) and  $n = 400$  (Tables 2, 3), respectively. For simplicity, we assume the same effect size for all of the regression coefficients. For each set of data, we perform the original and modified Lancaster procedures to assess the pathway data by combining  $p$ -values from individual tests. We set nominal error rate to be 0.05. The simulation is repeated 1000 times.

Due to LD, SNPs from the same gene are correlated. We first assess the Type I error rate of the test statistics by testing  $H_0$  :

**Table 2 | Type I error and power for independent Lancaster Procedure and five approximations to correlated Lancaster Procedures when sample size = 400 and linkage disequilibrium  $D = 0.20$ .**

	Independent Lancaster procedure	$T_A$	$T_B$	$T_C$	$T_D$	$T_E$
<b>CASE I</b>						
$\beta = 0$	<i>0.13</i>	0.051	0.052	0.051	0.051	0.051
$\beta = 0.4$	1	1	1	1	1	1
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE II</b>						
$\beta = 0$	<i>0.134</i>	0.05	0.051	0.05	0.05	0.05
$\beta = 0.4$	0.999	0.997	0.998	0.998	0.998	0.997
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE III</b>						
$\beta = 0$	<i>0.116</i>	0.045	0.048	0.045	0.045	0.045
$\beta = 0.4$	0.986	0.908	0.915	0.908	0.908	0.908
$\beta = 0.6$	1	0.998	0.998	0.998	0.998	0.998
<b>CASE IV</b>						
$\beta = 0$	<i>0.109</i>	0.046	0.047	0.046	0.046	0.046
$\beta = 0.4$	1	1	1	1	1	1
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE V</b>						
$\beta = 0$	<i>0.135</i>	0.04	0.043	0.041	0.041	0.041
$\beta = 0.4$	0.994	0.971	0.974	0.971	0.971	0.971
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE VI</b>						
$\beta = 0$	<i>0.135</i>	0.04	0.043	0.041	0.041	0.041
$\beta = 0.4$	0.986	0.939	0.948	0.939	0.939	0.939
$\beta = 0.6$	1	0.999	0.999	0.999	0.999	0.999

A Weight function is applied to adjust for the gene size\*.  
 \*The nominal error rate is set to be 0.05. Type I error rates are listed when  $\beta = 0$ . Power is listed when  $\beta > 0$ . Inflated Type I error rates are italicized.  
 \*A weight function  $w_i = 2/\sqrt{n_i}$  is applied to each test to adjust for the size of gene.

$\beta_1 = \dots = \beta_6 = 0$ . As shown in **Tables 1, 2**, the Type I error rate for the original Lancaster procedure is inflated ( $>0.05$ ) for all of the six cases. In contrast, five modified Lancaster procedures ( $T_A - T_E$ ) have well controlled Type I error rates ( $<0.05$ ).

The power of all test statistics was compared for regression coefficient values set at  $\beta = 0.4$  and  $\beta = 0.6$ , respectively. The results in **Tables 1, 2** suggest strong and comparable power among the modified Lancaster procedures. In most simulated cases, the proposed methods have more than 80% power to detect  $\beta = 0.4$ . When the effect size increases to  $\beta = 0.6$ , the power of proposed methods increases to 90% or above. Also the power of these tests improves as sample size increases from  $n = 200$  to  $n = 400$ .

We simulate different levels of LD for SNPs with  $D = 0, 1.5, 2$ , and uniform(0, maximum of LD). To save the space, we only show the results for  $D = 1.5$  (**Table 3**) and  $D = 2$  (**Tables 1, 2**). Our findings show that the inflation of Type I error rate for the original Lancaster procedure gets severe when LD is strong (**Tables 1, 2**). The modified Lancaster procedures ( $T_A - T_E$ ) have

well-controlled Type I error rates and power for both moderate and strong LD (**Tables 1–3**).

In **Table 4**, we assess the performance of all tests without a weighting function. We then compare the results in **Table 4** (without a weight function) vs. **Table 1** (with a weight function). All other simulation parameters are held the same in **Tables 1, 4**. We note that the original Lancaster procedure without a weighting function (**Table 4**) tends to have higher Type I error rates than the original Lancaster procedure with a weighting function (**Table 1**). For modified tests ( $T_A - T_E$ ), the power is increased when a weighting function is used. This confirms that an appropriate weight function is beneficial to the Lancaster procedure.

**CASE STUDY: RENAL TRANSPLANT TOLERANCE DATA**

We revisited a kidney transplant data first collected and analyzed by Newell et al. (2010). Data were downloaded from the GEO website with ID = GDS4266 (<http://www.ncbi.nlm.nih>).

**Table 3 | Type I error and power for independent Lancaster Procedure and five approximations to correlated Lancaster Procedures when sample size = 400 and linkage disequilibrium  $D = 0.15$ .**

	Independent Lancaster procedure	$T_A$	$T_B$	$T_C$	$T_D$	$T_E$
<b>CASE I</b>						
$\beta = 0$	<i>0.066</i>	0.043	0.045	0.043	0.044	0.043
$\beta = 0.4$	0.991	0.978	0.978	0.978	0.978	0.978
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE II</b>						
$\beta = 0$	<i>0.059</i>	0.031	0.035	0.031	0.031	0.031
$\beta = 0.4$	0.978	0.964	0.967	0.964	0.964	0.964
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE III</b>						
$\beta = 0$	<i>0.053</i>	0.029	0.034	0.029	0.03	0.029
$\beta = 0.4$	0.898	0.836	0.844	0.837	0.837	0.836
$\beta = 0.6$	0.999	0.996	0.997	0.996	0.996	0.996
<b>CASE IV</b>						
$\beta = 0$	<i>0.072</i>	0.041	0.045	0.041	0.041	0.041
$\beta = 0.4$	0.977	0.962	0.964	0.962	0.962	0.962
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE V</b>						
$\beta = 0$	<i>0.072</i>	0.041	0.045	0.041	0.041	0.041
$\beta = 0.4$	0.946	0.899	0.905	0.9	0.901	0.899
$\beta = 0.6$	0.999	0.996	0.996	0.996	0.996	0.996
<b>CASE VI</b>						
$\beta = 0$	<i>0.072</i>	0.041	0.045	0.041	0.041	0.041
$\beta = 0.4$	0.807	0.732	0.045	0.733	0.733	0.732
$\beta = 0.6$	0.978	0.965	0.045	0.965	0.965	0.965

A weight function is applied to adjust for the gene size\*.  
 \*The nominal error rate is set to be 0.05. Type I error rates are listed when  $\beta = 0$ . Power is listed when  $\beta > 0$ . Inflated Type I error rates are italicized.  
 \*A weight function  $w_i = 2/\sqrt{n_i}$  is applied to each test to adjust for the size of gene.

**Table 4 | Type I error and power for independent Lancaster Procedure and five approximations to correlated Lancaster Procedures when sample size = 200 and linkage disequilibrium  $D = 0.20$ .**

	Independent Lancaster procedure	$T_A$	$T_B$	$T_C$	$T_D$	$T_E$
<b>CASE I</b>						
$\beta = 0$	<i>0.106</i>	0.027	0.03	0.027	0.027	0.027
$\beta = 0.4$	1	0.997	0.997	0.997	0.997	0.997
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE II</b>						
$\beta = 0$	<i>0.1</i>	0.029	0.03	0.029	0.029	0.029
$\beta = 0.4$	0.935	0.801	0.812	0.801	0.803	0.801
$\beta = 0.6$	0.998	0.976	0.98	0.976	0.977	0.976
<b>CASE III</b>						
$\beta = 0$	<i>0.118</i>	0.041	0.042	0.041	0.041	0.041
$\beta = 0.4$	0.608	0.307	0.32	0.307	0.307	0.307
$\beta = 0.6$	0.881	0.663	0.679	0.665	0.666	0.663
<b>CASE IV</b>						
$\beta = 0$	<i>0.115</i>	0.037	0.04	0.038	0.038	0.037
$\beta = 0.4$	1	0.994	0.994	0.994	0.994	0.994
$\beta = 0.6$	1	1	1	1	1	1
<b>CASE V</b>						
$\beta = 0$	<i>0.115</i>	0.037	0.04	0.038	0.038	0.037
$\beta = 0.4$	0.78	0.487	0.5	0.488	0.489	0.487
$\beta = 0.6$	0.977	0.869	0.882	0.869	0.87	0.869
<b>CASE VI</b>						
$\beta = 0$	<i>0.115</i>	0.037	0.04	0.038	0.038	0.037
$\beta = 0.4$	0.782	0.579	0.589	0.579	0.58	0.579
$\beta = 0.6$	0.964	0.885	0.888	0.885	0.885	0.885

No Weight function is applied to adjust for the gene size\*.  
 \*The nominal error rate is set to be 0.05. Type I error rates are listed when  $\beta = 0$ . Power is listed when  $\beta > 0$ . Inflated Type I error rates are italicized.  
 \*These are the un-weighted tests with  $w_i = 2$  for all genes. We do not adjust the size of genes.

gov/sites/GDSbrowser?acc=GDS4266). A group of tolerant renal transplant recipients (Tolerant,  $n = 19$ ), as defined by stable graft function in the absence of immunosuppression for more than 1 year, were compared to subjects with stable graft function who were receiving standard immunotherapy (SI,  $n = 27$ ) as well as to a group of healthy controls (Control,  $n = 12$ ). Gene expression profiles of whole-blood total RNA from all subjects were measured by microarray. The goal of the study was to identify genetic variants associated with long-term allograft survival without the requirement for continuous immunosuppression, a condition known as allograft tolerance. Newell et al. (2010) performed statistical analysis to identify differentially expressed genes between the SI group and the Tolerant group. The results revealed a critical role for B cells in regulating alloimmunity, and provided a candidate set of genes for wider-scale screening of renal transplant recipients. However, no comprehensive pathway analysis was conducted by this group (Newell et al., 2010).

To further understand molecular mechanisms underlying renal allograft tolerance, we have applied the modified Lancaster

procedure to this dataset to identify candidate cellular pathways. Gene expression levels were normalized using Robust Multichip Average (rma) preprocessing methodology, which included background subtraction, quantile normalization, and summarization via median-polish.

Gene expression levels were summarized for a total of 54,675 probes from 21,049 genes. Expression levels were compared among three groups using the Bioconductor “Limma” package. Three pair wise comparisons were conducted, including: SI vs. Control, SI vs. Tolerant, and Tolerant vs. Control. Then three comparisons were combined into one  $F$ -test. This is equivalent to a One-Way ANOVA for each gene except that the residual mean squares have been moderated across genes.  $P$ -values from multiple hypothesis testing were adjusted by FDR (Benjamini and Hochberg, 1995). Our results of differentially expressed genes are consistent with the previous published work. See Newell et al. (2010) for the gene analysis findings.

Although (Newell et al., 2010) identified a set of differentially expressed genes, our analysis demonstrates that these significant

**Table 5 | Top 10 significant pathways detected by the modified Lancaster procedure ( $T_A$ ).**

GO accession	Pathway name	Gene ontology	URL	#Gene	#Probe	Adjusted $P$ -value
GO:0030183	B cell differentiation	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/B_CELL_DIFFERENTIATION">http://www.broadinstitute.org/gsea/msigdb/cards/B_CELL_DIFFERENTIATION</a>	12	29	0.003541
GO:0042113	B cell activation	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/B_CELL_ACTIVATION">http://www.broadinstitute.org/gsea/msigdb/cards/B_CELL_ACTIVATION</a>	20	45	0.003541
GO:0003823	Antigen binding	Molecular function	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/ANTIGEN_BINDING">http://www.broadinstitute.org/gsea/msigdb/cards/ANTIGEN_BINDING</a>	23	51	0.003541
GO:0004709	Map kinase kinase kinase activity	Molecular function	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/MAP_KINASE_KINASE_KINASE_ACTIVITY">http://www.broadinstitute.org/gsea/msigdb/cards/MAP_KINASE_KINASE_KINASE_ACTIVITY</a>	10	32	0.003541
GO:0017148	Negative regulation of translation	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/NEGATIVE_REGULATION_OF_TRANSLATION">http://www.broadinstitute.org/gsea/msigdb/cards/NEGATIVE_REGULATION_OF_TRANSLATION</a>	23	36	0.003541
GO:0042493	Response to drug	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/RESPONSE_TO_DRUG">http://www.broadinstitute.org/gsea/msigdb/cards/RESPONSE_TO_DRUG</a>	20	35	0.004669
GO:0001772	Immunological synapse	Cellular component	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/IMMUNOLOGICAL_SYNAPSE">http://www.broadinstitute.org/gsea/msigdb/cards/IMMUNOLOGICAL_SYNAPSE</a>	10	18	0.006603
GO:0030098	Lymphocyte differentiation	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/LYMPHOCYTE_DIFFERENTIATION">http://www.broadinstitute.org/gsea/msigdb/cards/LYMPHOCYTE_DIFFERENTIATION</a>	26	53	0.007986
GO:0042036	Negative regulation of cytokine biosynthetic process	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/NEGATIVE_REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS">http://www.broadinstitute.org/gsea/msigdb/cards/NEGATIVE_REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS</a>	12	21	0.008582
GO:0009890	Negative regulation of biosynthetic process	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/NEGATIVE_REGULATION_OF_BIOSYNTHETIC_PROCESS">http://www.broadinstitute.org/gsea/msigdb/cards/NEGATIVE_REGULATION_OF_BIOSYNTHETIC_PROCESS</a>	30	48	0.008582

genes have small effect sizes with fold changes  $<1.5$ . Therefore, a limited number of individual genes in the absence of a biological context is inadequate to explain the total variation of allograft tolerance among renal transplant patients.

To address this issue, we performed the modified Lancaster procedure ( $T_A$ ) as described in Section Correlated Lancaster Procedures to combine  $p$ -values from pathways. Combining  $p$ -values allows us to integrate small effects in pathway and gain the power of statistical testing. A total of 1454 Gene Ontology human pathway gene sets were analyzed. The size of pathways ranged from 9 genes to 2131 genes, with a median of 27 genes per pathway. Also, the number of probes per gene was highly variable. In order to map genes to pathways, we removed genes without gene symbols from the analysis. Among 21,049 genes with gene symbols, approximately 48% ( $n = 10161$ ) of genes were interrogated with a single probe, 26% ( $n = 5389$ ) of genes were queried using 2 probes, 14% ( $n = 2842$ ) of genes were assessed by 3 probes. There were 3 or more probes for each on the remaining genes (range: 4–17). This finding indicates that larger genes would have more  $p$ -values and a stronger impact to pathway analysis. To

prevent this bias, we set the weight function as  $w_i = 2/\sqrt{n_i}$  where  $n_i$  is the number of probes for the  $i$ th gene.

We performed pathway analysis for the One-Way ANOVA test and three pair wise comparisons. The top 10 significant pathways based on the One-Way ANOVA test are listed in **Table 5**. The top two pathways, B cell differentiation (GO:0030183) and B cell activation (GO:0042113), confirm the signature of B cell involvement described by Newell et al. (2010). Furthermore, we identified other pathways related to B cell activation and function. These include antigen binding (GO:0003823), map kinase kinase activity (GO:0004709) and lymphocyte differentiation (GO:0030098). These pathways are biologically consistent with the proposed role of B-lymphocytes in renal transplant tolerance reported by Newell et al. In contrast, when we performed the traditional Fisher's method without considering correlation structures (LD) within pathways or applying a weighting function to compensate for variability in the number of probes per gene, the result was a list of larger pathways, some containing  $>1000$  genes, describing more general cellular processes and not specifically related to immune functions (See **Table 6**, #gene and

**Table 6 | Top 10 significant pathways detected by the traditional Fisher's method.**

GO accession	Pathway name	Gene ontology	URL	# Gene	# Probes	Adjusted P-value
GO:0005737	Cytoplasm	Cellular component	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/CYTOPLASM">http://www.broadinstitute.org/gsea/msigdb/cards/CYTOPLASM</a>	2078	4986	0.E+00
GO:0005634	Nucleus	Cellular component	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/NUCLEUS">http://www.broadinstitute.org/gsea/msigdb/cards/NUCLEUS</a>	1393	3588	0.E+00
GO:0043283	Biopolymer metabolic process	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/BIOPOLYMER_METABOLIC_PROCESS">http://www.broadinstitute.org/gsea/msigdb/cards/BIOPOLYMER_METABOLIC_PROCESS</a>	1653	4240	0.E+00
GO:0016020	Membrane	Cellular component	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/MEMBRANE">http://www.broadinstitute.org/gsea/msigdb/cards/MEMBRANE</a>	1954	4395	3.E-307
GO:0006139	Nucleobase, nucleoside, nucleotide, and nucleic acid metabolic process	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/NUCLEOBASENUCLEOSIDENUCLEOTIDE_AND_NUCLEIC_ACID_METABOLIC_PROCESS">http://www.broadinstitute.org/gsea/msigdb/cards/NUCLEOBASENUCLEOSIDENUCLEOTIDE_AND_NUCLEIC_ACID_METABOLIC_PROCESS</a>	1217	3112	6.E-305
GO:0007165	Signal transduction	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/SIGNAL_TRANSDUCTION">http://www.broadinstitute.org/gsea/msigdb/cards/SIGNAL_TRANSDUCTION</a>	1604	3826	1.E-296
GO:0044425	Membrane part	Cellular component	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/MEMBRANE_PART">http://www.broadinstitute.org/gsea/msigdb/cards/MEMBRANE_PART</a>	1638	3670	4.E-251
GO:0019538	Protein metabolic process	Biological process	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/PROTEIN_METABOLIC_PROCESS">http://www.broadinstitute.org/gsea/msigdb/cards/PROTEIN_METABOLIC_PROCESS</a>	1205	3022	2.E-245
GO:0044422	Organelle part	Cellular component	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/ORGANELLE_PART">http://www.broadinstitute.org/gsea/msigdb/cards/ORGANELLE_PART</a>	1173	2934	1.E-230
GO:0044446	Intracellular organelle part	Cellular component	<a href="http://www.broadinstitute.org/gsea/msigdb/cards/INTRACELLULAR_ORGANELLE_PART">http://www.broadinstitute.org/gsea/msigdb/cards/INTRACELLULAR_ORGANELLE_PART</a>	1168	2923	4.E-230

#probe). Furthermore, when comparing the SI group and the Control group, the traditional method identified 1078 significant pathways while our proposed method narrowed the list down to 64 significant pathways (adjusted  $p$ -value  $<0.05$ ). The increase in number of significant pathways identified by the traditional approach is primarily due to false positive discovery, and is consistent with the inflation of Type I error rate as presented in Section Empirical Assessments. Thus, by accounting for correlation structures (LD) within pathways and the number of probes per gene, our proposed method minimized identification of larger, non-specific cellular processes pathways, and instead revealed more focused and functionally relevant biological pathways implicating a role for a humoral immune response in immunotolerance to renal transplants (See **Table 5**, #gene and #probe).

## DISCUSSION AND CONCLUSIONS

Modifications to the Lancaster procedure are proposed to take correlations among  $p$ -values into account. Extensive simulation studies show that the original Lancaster procedure has inflated Type I error rates due to correlation among  $p$ -values. By using permutation approach to estimate the correlation among  $p$ -values, the proposed methods have well-controlled Type I error rates and maintain strong power to detect signals related to SNPs in pathways.

Among five proposed approximation methods ( $T_A, \dots, T_E$ ), the Satterthwaite approximation ( $T_A$ ) is the most computationally efficient. Other approximation methods ( $T_B, \dots, T_E$ ) are based on the Satterthwaite approximation. Therefore, we recommend using the Satterthwaite approximation ( $T_A$ ) as the standard procedure to modify the Lancaster procedure. Among other approximation methods, simulation results in Section Correlated Lancaster Procedures show that, for data with stronger internal correlation,  $T_D$  and  $T_E$  have better approximation than  $T_B$  and  $T_C$ . Our simulation study and the case study further provide evidence that  $T_D$  tends to have slightly higher power than the Satterthwaite approximation  $T_A$ . The R code for five approximation is posted at <http://d.web.umkc.edu/daih/>.

## ACKNOWLEDGMENTS

We thank two reviewers for their constructive comments, which helped us improve the manuscript. This work was supported in part by NSF grant DMS-1209112 (to Yuehua Cui).

## REFERENCES

- Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Stat.* 38, 303–324. doi: 10.1214/aoms/1177698949
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–333. doi: 10.2307/2346101
- Benjamini, Y., and Hochberg, Y. (1997). Multiple hypothesis testing with weights. *Scand. J. Stat.* 24, 407–417. doi: 10.1111/1467-9469.00072
- Cheng, C., and Pounds, S. (2007). False discovery rate paradigms for statistical analyses of microarray gene expression data. *Bioinformatics* 1, 436–446. doi: 10.6026/97320630001436
- Cui, Y., Kang, G., Sun, K., Qian, M., Romero, R., and Fu, W. (2008). Gene-centric genomewide association study via entropy. *Genetics* 179, 637–650. doi: 10.1534/genetics.107.082370
- Dai, H., Bhandary, M., Becker, M. L., Leeder, S. J., Gaedigk, R., and Motsinger-Reif, A. A. (2012a). Global tests of  $p$ -values for multifactor dimensionality reduction models in selection of optimal number of target genes. *BioData Min.* 5:3. doi: 10.1186/1756-0381-5-3
- Dai, H., Charnigo, R., Srivastava, T., Talebizadeh, Z., and Ye, S. (2012b). Integrating  $P$ -values for genetic and genomic data analysis. *J. Biom. Biostat.* 3:e117. doi: 10.4172/2155-6180.1000e117
- Dai, H., and Charnigo, R. (2008). Omnibus testing and gene filtration in microarray data analysis. *J. Appl. Stat.* 35, 31–47. doi: 10.1080/02664760701683528
- Fisher, R. A. (1922). On the interpretation of  $\chi^2$  from contingency tables and calculation of  $p$ . *J. R. Stat. Soc. A* 85, 87–94. doi: 10.2307/2340521
- Fisher, R. A. (1932). *Statistical Methods for Research Workers*. London: Oliver & Boyd.
- Fisher, R. A., and Cornish, E. A. (1960). The percentile points of distributions having known cumulants. *Technometrics* 2, 209–225. doi: 10.1080/00401706.1960.10489895
- Genovese, C. R., Roeder, K., and Wasserman, L. (2006). False discovery control with  $p$ -value weighting. *Biometrika* 93, 509–524. doi: 10.1093/biomet/93.3.509
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280. doi: 10.1093/nar/gkh063
- Khatri, P., Sirota, M., and Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.* 8:e1002375. doi: 10.1371/journal.pcbi.1002375
- Koziol, J. A. (1996). A note on Lancaster's procedure for the combination of independent events. *Biom. J.* 38, 653–660. doi: 10.1002/bimj.4710380603
- Lancaster, H. D. (1961). The combination of probabilities: an application of orthonormal functions. *Aust. J. Stat.* 3, 20–33. doi: 10.1111/j.1467-842X.1961.tb00058.x
- Li, S., Williams, B. L., and Cui, Y. (2011). A combined  $p$ -value approach to infer pathway regulations in eQTL mapping. *Stat. Interface* 4, 389–402. doi: 10.4310/SII.2011.v4.n3.a13
- Littell, R. C., and Folks, J. L. (1971). Asymptotic optimality of Fisher's method of combining independent tests. *J. Am. Stat. Assoc.* 66, 802–806. doi: 10.1080/01621459.1971.10482347
- Littell, R. C., and Folks, J. L. (1973). Asymptotic optimality of Fisher's method of combining independent tests. II. *J. Am. Stat. Assoc.* 68, 193–194. doi: 10.1080/01621459.1973.10481362
- Lugannani, R., and Rice, S. O. (1980). Saddlepoint approximation for the sum of independent random variables. *Adv. Appl. Probab.* 12, 475–490. doi: 10.2307/1426607
- Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., et al. (2009). Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 37, D619–D622. doi: 10.1093/nar/gkn863
- Mosteller, F., and Bush, R. R. (1954). Selected quantitative technique. *Handb. Soc. Psychol.* 1, 289–334.
- Newell, K. A., Asare, A., Kirk, A. D., Gisler, T. D., Bourcier, K., Suthanthiran, M., et al. (2010). Identification of a B cell signature associated with renal transplant tolerance in humans. *J. Clin. Invest.* 120, 1836–1847. doi: 10.1172/JCI39933
- Nikolsky, Y., and Bryant, J. (2009). Protein networks and pathway analysis. Preface. *Methods Mol. Biol.* 563, v–vii. doi: 10.1007/978-1-60761-175-2
- Patnaik, P. B. (1949). The non-central  $\chi^2$  - and  $F$  - distributions and their applications. *Biometrika* 36, 202–232.
- Peng, G., Luo, L., Siu, H., Zhu, Y., Hu, P., Hong, S., et al. (2010). Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur. J. Hum. Genet.* 18, 111–117. doi: 10.1038/ejhg.2009.115
- Stouffer, S., DeVinney, L., and Suchman, E. (1949). *The American Soldier: Adjustment during Army Life*. Vol. 1. Princeton, NJ: Princeton University Press.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., and Rothman, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J. Natl. Cancer Inst.* 96, 434–442. doi: 10.1093/jnci/djh075
- Wang, K., Li, M., and Bucan, M. (2007). Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* 81, 1278–1283. doi: 10.1086/522374
- Wilson, E. B., and Hilferty, M. M. (1931). The distribution of chi-square. *Proc. Natl. Acad. Sci. U.S.A.* 17, 684–688. doi: 10.1073/pnas.17.12.684

Wu, M. C., and Lin, X. (2009). Prior biological knowledge-based approaches for the analysis of genome-wide expression profiles using gene sets and pathways. *Stat. Methods Med. Res.* 18, 577–593. doi: 10.1177/0962280209351925

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 05 November 2013; accepted: 27 January 2014; published online: 20 February 2014.

*Citation:* Dai H, Leeder JS and Cui Y (2014) A modified generalized Fisher method for combining probabilities from dependent tests. *Front. Genet.* 5:32. doi: 10.3389/fgene.2014.00032

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Dai, Leeder and Cui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.