

# A General Model for Multilocus Epistatic Interactions in Case-Control Studies

Zhong Wang<sup>1,2</sup>, Tian Liu<sup>3</sup>, Zhenwu Lin<sup>4</sup>, John Hegarty<sup>4</sup>, Walter A. Koltun<sup>4</sup>, Rongling Wu<sup>1,2\*</sup>

**1** Center for Statistical Genetics, Pennsylvania State University, Hershey, Pennsylvania, United States of America, **2** Pennsylvania State Cancer Institute, Pennsylvania State University, Hershey, Pennsylvania, United States of America, **3** Human Genetics Group, Genome Institute of Singapore, Singapore, Singapore, **4** Department of Surgery, Pennsylvania State University, Hershey, Pennsylvania, United States of America

## Abstract

**Background:** Epistasis, i.e., the interaction of alleles at different loci, is thought to play a central role in the formation and progression of complex diseases. The complexity of disease expression should arise from a complex network of epistatic interactions involving multiple genes.

**Methodology:** We develop a general model for testing high-order epistatic interactions for a complex disease in a case-control study. We incorporate the quantitative genetic theory of high-order epistasis into the setting of cases and controls sampled from a natural population. The new model allows the identification and testing of epistasis and its various genetic components.

**Conclusions:** Simulation studies were used to examine the power and false positive rates of the model under different sampling strategies. The model was used to detect epistasis in a case-control study of inflammatory bowel disease, in which five SNPs at a candidate gene were typed, leading to the identification of a significant three-locus epistasis.

**Citation:** Wang Z, Liu T, Lin Z, Hegarty J, Koltun WA, et al. (2010) A General Model for Multilocus Epistatic Interactions in Case-Control Studies. PLoS ONE 5(8): e11384. doi:10.1371/journal.pone.0011384

**Editor:** Zvia Agur, Institute for Medical Biomathematics, Israel

**Received:** January 15, 2010; **Accepted:** May 31, 2010; **Published:** August 18, 2010

**Copyright:** © 2010 Wang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work is supported by National Science Foundation (NSF) grant DMS/NIGMS-0540745 and the Changjiang Scholars Award. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: rwu@hes.hmc.psu.edu

## Introduction

The complexity of biological systems arises from the highly interactive relationships of their components [1,2]. Thus, it is likely that the metabolic pathways for a phenotypic trait or disease involve multiple interacting gene products and regulatory loci that could generate a complex network of genetic actions and interactions [3,4]. Current genome-wide linkage or association studies have been able to detect genetic actions of individual genes involved in the phenotypic diversity of a complex trait [5–12]. Given its ubiquitousness in controlling complex traits and diseases, epistasis resulting from interactions between alleles at different genes has now received increasing attention in genetic studies [13,14]. However, many of these studies focus on the identification of low-order pairwise epistasis, leaving epistatic interactions of high orders, their frequency and impact on genetic variation, unexplored.

More recently, Stich et al. [15] developed a linkage mapping approach to uncover three-way interactions among different quantitative trait loci (QTLs) using a mating design. Beerenwinkel et al. [16] proposed a mathematical approach for describing multi-way genetic interactions and employing it to study the genetic structure of fitness landscapes for *Escherichia coli*. Based on the analysis of pathway fragments, Imielinski and Belta [17] used a genome-scale knockout design to detect high-order epistatic relationships between components of large metabolic networks. Hansen and Wagner [18] showed that higher-order genetic

interactions are potentially important if the total genomic mutation rate is large and the interaction density among loci is not too low. With the widespread availability of high-throughput genotyping technology, there is a pressing need to estimate higher-order epistasis involving any number of genes and assess the role of epistasis in the creation and maintenance of genetic variation for complex traits.

The motivation of this study is to develop a general model for estimating epistasis of any order with multilocus single nucleotide polymorphism (SNP) data in case-control studies. In particular, the model allows the estimation and testing of high-order epistasis. Because of its easy sample collection, a population-based case-control design has been widely used in candidate gene or genome-wide association studies [19–21]. By comparing genotype frequencies for a gene in unrelated individuals with the disease and healthy controls, this design has power to test the significance of the association between the gene and disease. However, only a few studies used a case-control design to characterize epistasis [19] and, also, the epistasis they defined on the basis of logistic regression models presents a computational complexity. The new model described in this article has, for the first time, embedded quantitative genetic principles into a chi-square test framework, allowing the dissection of overall multilocus genetic effects into various components including epistatic interactions of high orders. The model was validated through simulation studies and a real data analysis.

### Large Quantitative Genetic Models for Epistasis

Epistasis was originally defined as the expression of an allele at one locus masked by an allele at another locus [22]. This concept was then explained in a statistical manner by Fisher [23] as the deviation of genetic action from additivity in a linear model. Fisher’s definition allows epistasis to be quantified in different forms based on its biological meaning determined by Bateson [22]. For a two-locus epistasis, all possible forms of epistasis include the interactions between additive effects at the two loci, additive effect at the first locus and dominant effect at the second locus, dominant effect at the first locus and additive effect at the second locus, and dominant effects at the two loci. Each of these epistatic forms contributes differently to the overall genetic value of a two-locus genotype. We used Mather and Jinks’ formulation [24] to partition a genotypic value into its different components including epistasis.

### Two-locus Epistasis

Suppose there are two loci, **A** with two alleles *A* and *a* and **B** with two alleles *B* and *b*, which form nine two-locus genotypes. Let  $\mu_{j_1j_2}$  denote the genetic value of an arbitrary genotype  $j_1j_2$  ( $j_1 = 2, 1, 0$  for genotypes *AA*, *Aa*, and *aa*;  $j_2 = 2, 1, 0$  for genotypes *BB*, *Bb*, and *bb*, respectively). We dissect  $\mu_{j_1j_2}$  into different components as table 1.

Where  $\mu$  is the overall mean,  $a_1$  and  $a_2$  are the additive effect at genes **A** and **B**,  $d_1$  and  $d_2$  are the dominant effect at genes **A** and **B**, respectively, and  $i_{aa}$ ,  $i_{ad}$ ,  $i_{da}$ , and  $i_{dd}$  are the additive  $\times$  additive, additive  $\times$  dominant, dominant  $\times$  additive, and dominant  $\times$  dominant epistatic interactions between the two genes, respectively.

The dissection of genotypic values is expressed, in matrix form, as

$$\begin{pmatrix} \mu_{00} \\ \mu_{01} \\ \mu_{02} \\ \mu_{10} \\ \mu_{11} \\ \mu_{12} \\ \mu_{20} \\ \mu_{21} \\ \mu_{22} \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} \mu_0 \\ a_1 \\ a_2 \\ d_1 \\ d_2 \\ i_{aa} \\ i_{ad} \\ i_{da} \\ i_{dd} \end{pmatrix} \quad (1)$$

The genetic effect parameters can be solved using

$$\begin{pmatrix} \mu_0 \\ a_1 \\ a_2 \\ d_1 \\ d_2 \\ i_{aa} \\ i_{ad} \\ i_{da} \\ i_{dd} \end{pmatrix} = \frac{1}{2^2} \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 1 \\ -1 & 0 & -1 & 2 & 0 & 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 & -1 & 2 & -1 \\ 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 \\ 1 & -2 & 1 & 0 & 0 & 0 & -1 & 2 & -1 \\ 1 & 0 & -1 & -2 & 0 & 2 & 1 & 0 & -1 \\ 1 & -2 & 1 & -2 & 4 & -2 & 1 & -2 & 1 \end{pmatrix} \times \begin{pmatrix} \mu_{00} \\ \mu_{01} \\ \mu_{02} \\ \mu_{10} \\ \mu_{11} \\ \mu_{12} \\ \mu_{20} \\ \mu_{21} \\ \mu_{22} \end{pmatrix} \quad (2)$$

### Three-locus Epistasis

Adding a locus, **C** with two alleles *C* and *c*, to the two-gene model generates 27 three-locus genotypes, expressed as  $j_1j_2j_3$  ( $j_3 = 2, 1, 0$  for genotypes *CC*, *Cc*, and *cc*, respectively). A three-genotypic value ( $\mu_{j_1j_2j_3}$ ) is dissected into the following components:

- (1) the overall mean  $\mu$ ;
- (2) the main genetic effects including the three additive effects ( $a_1$ ,  $a_2$ , and  $a_3$ ) at genes **A**, **B**, and **C**, and the three dominant effects ( $d_1$ ,  $d_2$ , and  $d_3$ ) at genes **A**, **B**, and **C**;
- (3) the two-way interaction effects including the additive  $\times$  additive ( $i_{a_1a_2}$ ), additive  $\times$  dominant ( $i_{a_1d_2}$ ), dominant  $\times$  additive ( $i_{d_1a_2}$ ), and dominant  $\times$  dominant ( $i_{d_1d_2}$ ) epistasis between genes **A** and **B**, the additive  $\times$  additive ( $i_{a_1a_3}$ ), additive  $\times$  dominant ( $i_{a_1d_3}$ ), dominant  $\times$  additive ( $i_{d_1a_3}$ ), and dominant  $\times$  dominant ( $i_{d_1d_3}$ ) epistasis between genes **A** and **C**, and additive  $\times$  additive ( $i_{a_2a_3}$ ), additive  $\times$  dominant ( $i_{a_2d_3}$ ), dominant  $\times$  additive ( $i_{d_2a_3}$ ), and dominant  $\times$  dominant ( $i_{d_2d_3}$ ) epistasis between genes **B** and **C**;
- (4) the three-way interaction effects including the additive  $\times$  additive  $\times$  additive ( $i_{a_1a_2a_3}$ ), additive  $\times$  additive  $\times$  dominant ( $i_{a_1a_2d_3}$ ), additive  $\times$  dominant  $\times$  additive ( $i_{a_1d_2a_3}$ ), dominant  $\times$  additive  $\times$  additive ( $i_{d_1a_2a_3}$ ), additive  $\times$  dominant  $\times$  dominant ( $i_{a_1d_2d_3}$ ), dominant  $\times$  additive  $\times$  dominant ( $i_{d_1a_2d_3}$ ), dominant  $\times$  dominant  $\times$  additive ( $i_{d_1d_2a_3}$ ), and dominant  $\times$  dominant  $\times$  dominant ( $i_{d_1d_2d_3}$ ) epistasis among genes **A**, **B**, and **C**.

Mather and Jinks’ theory is used to formulate the relationships between genotypic values and genetic effects, expressed as

**Table 1.** The genetic effect components of two-locus genotypes.

Genotype	Value	Component								
		$\mu$	$a_1$	$a_2$	$d_1$	$d_2$	$i_{aa}$	$i_{ad}$	$i_{da}$	$i_{dd}$
<i>AABB</i>	$\mu_{22}$	+	+	+			+			
<i>AABb</i>	$\mu_{21}$	+	+			+		+		
<i>AAbb</i>	$\mu_{20}$	+	+	-				-		
<i>AaBB</i>	$\mu_{12}$	+		+	+				+	
<i>AaBb</i>	$\mu_{11}$	+			+	+				+
<i>Aabb</i>	$\mu_{10}$	+		-	+				-	
<i>aaBB</i>	$\mu_{02}$	+	-	+				-		
<i>aaBb</i>	$\mu_{01}$	+	-			+		-		
<i>aabb</i>	$\mu_{00}$	+	-	-				+		

doi:10.1371/journal.pone.0011384.t001



- If  $e_1 = e_2 = \dots = e_n = \phi$ , then  $e_{\phi, \phi, \dots, \phi} = \mu$ ;
- If  $e_1 = a_1, e_2 = e_3 = \dots = e_n = \phi$ , then  $e_{a_1, \phi, \dots, \phi} = a_1$ ;
- ...;
- If  $e_1 = a_2, e_2 = a_2, e_3 = e_4 = \dots = e_n = \phi$ , then  $e_{a_1, a_2, \phi, \dots, \phi} = i_{a_1 a_2}$ ;
- ...;
- If  $e_1 = a_1, e_2 = a_2, \dots, e_n = a_n$ , then  $e_{a_1, a_2, \dots, a_n} = i_{a_1 a_2 \dots a_n}$ .
- ...;
- If  $e_1 = d_1, e_2 = d_2, \dots, e_n = d_n$ , then  $e_{d_1, d_2, \dots, d_n} = i_{d_1 d_2 \dots d_n}$ .

By letting  $\phi_l = 0, a_l = 1$  and  $d_l = 2$  ( $l = 1, \dots, n$ ), we express the value of a general multi-locus genotype as

$$\mu_{j_1 j_2 \dots j_n} = \underbrace{\sum_{e_n=0}^2 \dots \sum_{e_1=0}^2}_{n} [f_{\mu}(j_1, j_2, \dots, j_n, e_1, e_2, \dots, e_n) \cdot e_{e_1, e_2, \dots, e_n}] \quad (5)$$

where

$$f_{\mu}(j_1, j_2, \dots, j_n, e_1, e_2, \dots, e_n) = \prod_{l=1}^n g_{\mu}(j_l, e_l), \quad (6)$$

with

$$g_{\mu}(j_l, e_l) = \begin{vmatrix} L(j=0) & L(j=1) & L(j=2) \\ \times \begin{vmatrix} 1 & (-1) & 0 \\ 1 & 0 & (-1) \\ 1 & 0 & 0 \end{vmatrix} & \times \begin{vmatrix} L(e_l = \phi) \\ L(e_l = a_l) \\ L(e_l = d_l) \end{vmatrix} \end{vmatrix} \quad (7)$$

$L$  is a logical judgment function that can return 1 if the condition is true otherwise return 0.

The genetic effect parameters can be estimated by solving the linear equations using

$$e_{e_1, e_2, \dots, e_n} = \underbrace{\sum_{j_n=0}^2 \dots \sum_{j_1=0}^2}_{n} [f_i(e_1, e_2, \dots, e_n, j_1, j_2, \dots, j_n) \cdot \mu_{j_1 j_2 \dots j_n}] \quad (8)$$

where

$$f_i(e_1, e_2, \dots, e_n, j_1, j_2, \dots, j_n) = \frac{1}{2^n} \prod_{l=1}^n g_i(e_l, j_l) \quad (9)$$

with

$$g_i(e_l, j_l) = \begin{vmatrix} L(j=0) & L(j=1) & L(j=2) \\ \times \begin{vmatrix} 1 & (-1) & (-1) \\ 0 & 0 & 2 \\ 1 & 1 & (-1) \end{vmatrix} & \times \begin{vmatrix} L(e_l = \phi) \\ L(e_l = a_l) \\ L(e_l = d_l) \end{vmatrix} \end{vmatrix} \quad (10)$$

Equation (8) gives a general form for main and interaction genetic effects among an arbitrary number of loci. Mathematical algorithms for solving epistatic equations are given in Text S1.

### Testing Epistasis

Based on the definitions, we now provide a procedure for testing epistasis of different kinds and orders with multilocus genetic data. Consider a case-control study in which  $m$  cases (there is a disease) and  $n$  controls (there is no disease) are selected randomly from a natural population. Case and control groups are matched for

demographical factors such as age, race, gender, life style, and body mass. All subjects from the case and control groups are genotyped genome-wide or for particular chromosomal regions of interest, depending on the purpose of the study. Let  $m_{j_1 j_2 j_3}^{(1)}$  and  $m_{j_1 j_2 j_3}^{(2)}$  denote the observations of a general genotype  $j_1 j_2 j_3$  ( $j_1, j_2, j_3 = 2, 1, 0$ ) derived from three markers **A**, **B**, and **C**. Based on Mather and Jinks' partition of genotypic values [24], we calculate genetic effect parameters from genotypic values using equation (4). For both cases and controls, the genotypic values used to calculate each effect parameter are dissolved into two groups, plus and minus, which forms a 2 (cases and control)  $\times$  2 (plus and minus) contingency table. For example, the contingency table for testing the additive  $\times$  additive  $\times$  additive epistatic effect is expressed as table 2.

From the table, the  $\chi^2$  test statistic is calculated and compared with the critical threshold with one degree of freedom. We proved that the test statistics under the null hypothesis calculated from the above contingency table follows a  $\chi^2$  distribution with less than one degree of freedom [25].

The contingency tables for testing the other parameters can be made similarly. For a particular group  $k$  ( $k = 1$  for cases, 2 for controls), the genotypic values used to calculate the three-way epistatic effect parameters are tabulated as table 3.

The thresholds for testing each of these three-locus epistases are derived, which are  $\chi_{0.05}^2 = 3.84, 3.20, 3.20, 3.20, 2.60, 2.60, 2.60$ , and 2.14, respectively. The genotypic values used to calculate the two-way epistatic effect parameters are tabulated as table 4.

The thresholds for testing each of these two-locus epistases are derived, which are  $\chi_{0.05}^2 = 3.84, 3.84, 3.84, 2.50, 2.50, 2.50, 3.20, 3.20, 3.20, 3.20$  and 3.20, respectively. The genotypic values used to calculate the main genetic effect parameters are tabulated as table 5.

The thresholds for testing each of these two-locus epistases are derived, which are  $\chi_{0.05}^2 = 3.84, 3.84, 3.84, 2.60, 2.60$  and 2.60, respectively. For an arbitrary number of markers, the genotypic values used to calculate the main and epistatic (of different orders) genetic effect parameters can be similarly divided into plus and minus groups, from which the  $\chi^2$  test statistics are calculated.

## Results

The model was used to analyze a case-control study aimed to detect genetic variants for inflammatory bowel disease (IBD) with candidate gene approaches [26]. As a member of the membrane associated guanylate kinase family, TDiscs large homolog (DLG5) plays a central role in maintaining cell junctions and cell shape and in clustering channel proteins at the cell surface [27]. Five single nucleotide polymorphisms (SNPs), Arg30Gln, Glu514Gln, Pro979Leu, Gly1066Gly, and Pro1371Gln, genotyped at DLG5 for both cases and controls are hoped to be associated with IBD. The cases include 115 sporadic IBD patients, aged from 22 to 66 years old, from the Milton S Hershey Medical Center, whereas the controls are 172 unrelated healthy individuals, aged from 15 to 81

**Table 2.** The  $\chi^2$  test statistics for the additive  $\times$  additive  $\times$  additive epistatic effect.

	Plus	Minus
Cases	$m_{222}^{(1)} + m_{200}^{(1)} + m_{020}^{(1)} + m_{002}^{(1)}$	$m_{220}^{(1)} + m_{202}^{(1)} + m_{022}^{(1)} + m_{000}^{(1)}$
Controls	$m_{222}^{(2)} + m_{200}^{(2)} + m_{020}^{(2)} + m_{002}^{(2)}$	$m_{220}^{(2)} + m_{202}^{(2)} + m_{022}^{(2)} + m_{000}^{(2)}$

doi:10.1371/journal.pone.0011384.t002

**Table 3.** The  $\chi^2$  test statistics for the three-way epistatic effect parameters.

Parameter	Plus	Minus
$i_{a_1 a_2 a_3}$	$m_{222}^{(k)} + m_{200}^{(k)} + m_{020}^{(k)} + m_{002}^{(k)}$	$m_{220}^{(k)} + m_{202}^{(k)} + m_{022}^{(k)} + m_{000}^{(k)}$
$i_{a_1 a_2 d_3}$	$2m_{221}^{(k)} + m_{202}^{(k)} + m_{200}^{(k)} + m_{022}^{(k)} + m_{020}^{(k)} + 2m_{001}^{(k)}$	$m_{222}^{(k)} + m_{220}^{(k)} + 2m_{201}^{(k)} + 2m_{021}^{(k)} + m_{002}^{(k)} + m_{000}^{(k)}$
$i_{a_1 d_2 a_3}$	$m_{220}^{(k)} + 2m_{212}^{(k)} + m_{200}^{(k)} + m_{022}^{(k)} + 2m_{010}^{(k)} + m_{002}^{(k)}$	$m_{222}^{(k)} + 2m_{210}^{(k)} + m_{202}^{(k)} + m_{020}^{(k)} + 2m_{012}^{(k)} + m_{000}^{(k)}$
$i_{a_1 d_2 d_3}$	$m_{222}^{(k)} + m_{220}^{(k)} + 4m_{211}^{(k)} + m_{202}^{(k)} + m_{200}^{(k)} + 2m_{201}^{(k)} + 2m_{012}^{(k)} + 2m_{010}^{(k)} + 2m_{001}^{(k)}$	$2m_{221}^{(k)} + 2m_{212}^{(k)} + 2m_{210}^{(k)} + 2m_{201}^{(k)} + m_{022}^{(k)} + m_{020}^{(k)} + 4m_{011}^{(k)} + m_{002}^{(k)} + m_{000}^{(k)}$
$i_{d_1 a_2 a_3}$	$m_{220}^{(k)} + m_{202}^{(k)} + 2m_{122}^{(k)} + 2m_{100}^{(k)} + m_{020}^{(k)} + m_{002}^{(k)}$	$m_{222}^{(k)} + m_{200}^{(k)} + 2m_{120}^{(k)} + 2m_{102}^{(k)} + m_{022}^{(k)} + m_{000}^{(k)}$
$i_{d_1 a_2 d_3}$	$m_{222}^{(k)} + m_{220}^{(k)} + 2m_{201}^{(k)} + 4m_{121}^{(k)} + 2m_{102}^{(k)} + 2m_{100}^{(k)} + m_{022}^{(k)} + m_{020}^{(k)} + 2m_{001}^{(k)}$	$2m_{221}^{(k)} + m_{202}^{(k)} + m_{200}^{(k)} + 2m_{122}^{(k)} + 2m_{120}^{(k)} + 4m_{101}^{(k)} + 2m_{021}^{(k)} + m_{002}^{(k)} + m_{000}^{(k)}$
$i_{d_1 d_2 a_3}$	$m_{222}^{(k)} + 2m_{210}^{(k)} + m_{202}^{(k)} + 2m_{120}^{(k)} + 4m_{112}^{(k)} + 2m_{100}^{(k)} + m_{022}^{(k)} + 2m_{010}^{(k)} + m_{002}^{(k)}$	$m_{220}^{(k)} + 2m_{212}^{(k)} + m_{200}^{(k)} + 2m_{122}^{(k)} + 4m_{110}^{(k)} + 2m_{102}^{(k)} + m_{020}^{(k)} + 2m_{012}^{(k)} + m_{000}^{(k)}$
$i_{d_1 d_2 d_3}$	$2m_{221}^{(k)} + 2m_{212}^{(k)} + 2m_{210}^{(k)} + 2m_{201}^{(k)} + 2m_{122}^{(k)} + 2m_{120}^{(k)} + 8m_{111}^{(k)} + 2m_{102}^{(k)} + 2m_{100}^{(k)} + 2m_{021}^{(k)} + 2m_{012}^{(k)} + 2m_{001}^{(k)}$	$m_{222}^{(k)} + m_{220}^{(k)} + 4m_{121}^{(k)} + 4m_{112}^{(k)} + 4m_{110}^{(k)} + 4m_{101}^{(k)} + m_{022}^{(k)} + m_{020}^{(k)} + 4m_{011}^{(k)} + m_{002}^{(k)} + m_{000}^{(k)}$

doi:10.1371/journal.pone.0011384.t003

years, from the Milton S Hershey Medical Center and Philadelphia gift of Life Donor Program. All the human tissues used for pathological studies and genetic analysis were approved by the Human Subjects Protection Offices of The Pennsylvania State University College of Medicine, and were undertaken with the understanding and written consent of each subject.

Because of a modest sample size used, our analysis will focus on a three-SNP analysis, although the model can deal with any number of SNPs. None of the five SNPs displays an additive genetic effect, but Arg30Gln, Pro979Leu, and Gly1066Gly were each found to trigger a significant dominant effect on the disease ( $p = 0.04 - 0.0002$ ) (table 6). There are 10 possible pairs for the five SNPs, with each pair subject to a two-locus epistatic analysis. The number and distribution of two-locus epistasis are given in table 7. It is interesting to see that significant two-locus epistasis was observed only between Arg30Gln and other SNPs including Pro979Leu with a significant main dominant effect and two non-significant SNPs (Glu514Gln and Pro1371Gln). The form of significant epistasis is limited to the interactions between the dominant effect at Arg30Gln and the additive/dominant effects at the other SNPs.

**Table 4.** The  $\chi^2$  test statistics for the two-way epistatic effect parameters.

Parameter	Plus	Minus
$i_{a_1 a_2}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$	$m_{020}^{(k)} + m_{022}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)}$
$i_{a_2 a_3}$	$m_{000}^{(k)} + m_{022}^{(k)} + m_{200}^{(k)} + m_{222}^{(k)}$	$m_{002}^{(k)} + m_{020}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)}$
$i_{a_1 a_3}$	$m_{000}^{(k)} + m_{020}^{(k)} + m_{202}^{(k)} + m_{222}^{(k)}$	$m_{002}^{(k)} + m_{022}^{(k)} + m_{200}^{(k)} + m_{220}^{(k)}$
$i_{d_1 d_2}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + 4m_{110}^{(k)} + 4m_{112}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$	$2m_{010}^{(k)} + 2m_{012}^{(k)} + 2m_{100}^{(k)} + 2m_{102}^{(k)} + 2m_{120}^{(k)} + 2m_{122}^{(k)} + 2m_{210}^{(k)} + 2m_{212}^{(k)}$
$i_{d_2 d_3}$	$m_{000}^{(k)} + m_{002}^{(k)} + 4m_{011}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + 4m_{211}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$	$2m_{001}^{(k)} + 2m_{010}^{(k)} + 2m_{012}^{(k)} + 2m_{021}^{(k)} + 2m_{020}^{(k)} + 2m_{210}^{(k)} + 2m_{212}^{(k)} + 2m_{221}^{(k)}$
$i_{d_1 d_3}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + 4m_{101}^{(k)} + 4m_{121}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$	$2m_{001}^{(k)} + 2m_{021}^{(k)} + 2m_{100}^{(k)} + 2m_{102}^{(k)} + 2m_{120}^{(k)} + 2m_{122}^{(k)} + 2m_{201}^{(k)} + 2m_{221}^{(k)}$
$i_{a_1 d_2}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + 2m_{210}^{(k)} + 2m_{212}^{(k)}$	$2m_{010}^{(k)} + 2m_{012}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$
$i_{a_1 d_3}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + 2m_{201}^{(k)} + 2m_{221}^{(k)}$	$2m_{001}^{(k)} + 2m_{021}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$
$i_{d_1 a_2}$	$m_{000}^{(k)} + m_{002}^{(k)} + 2m_{120}^{(k)} + 2m_{122}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)}$	$m_{020}^{(k)} + m_{022}^{(k)} + 2m_{100}^{(k)} + 2m_{102}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$
$i_{d_1 a_3}$	$m_{000}^{(k)} + m_{020}^{(k)} + 2m_{102}^{(k)} + 2m_{122}^{(k)} + m_{200}^{(k)} + m_{220}^{(k)}$	$m_{002}^{(k)} + m_{022}^{(k)} + 2m_{100}^{(k)} + 2m_{120}^{(k)} + m_{202}^{(k)} + m_{222}^{(k)}$
$i_{a_2 d_3}$	$m_{000}^{(k)} + m_{002}^{(k)} + 2m_{021}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + 2m_{221}^{(k)}$	$2m_{001}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + 2m_{201}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$
$i_{d_2 a_3}$	$m_{000}^{(k)} + 2m_{012}^{(k)} + m_{020}^{(k)} + m_{200}^{(k)} + 2m_{212}^{(k)} + m_{220}^{(k)}$	$m_{002}^{(k)} + 2m_{010}^{(k)} + m_{022}^{(k)} + m_{202}^{(k)} + 2m_{210}^{(k)} + m_{222}^{(k)}$

doi:10.1371/journal.pone.0011384.t004

The five SNPs produce 10 three-locus combinations which were analyzed by a three-locus epistasis model. Each combination has eight forms of three-SNP epistasis. Table 8 lists the test statistics for all possible combinations and forms of epistasis, with significant epistasis highlighted in boldface. The interactions among the additive effects at any three of the five SNPs were not significant; the same was also observed for the three-way dominant interactions. The significant three-locus epistasis must include both the additive and dominant effect at three SNPs. In general, Arg30Gln have more significant three-locus interactions and

**Table 5.** The  $\chi^2$  test statistics for the main epistatic effect parameters.

Parameter	Plus	Minus
$a_1$	$m_{200}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)}$
$a_2$	$m_{020}^{(k)} + m_{022}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)}$
$a_3$	$m_{002}^{(k)} + m_{022}^{(k)} + m_{202}^{(k)} + m_{222}^{(k)}$	$m_{000}^{(k)} + m_{020}^{(k)} + m_{200}^{(k)} + m_{220}^{(k)}$
$d_1$	$2m_{100}^{(k)} + 2m_{102}^{(k)} + 2m_{120}^{(k)} + 2m_{122}^{(k)}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$
$d_2$	$2m_{010}^{(k)} + 2m_{012}^{(k)} + 2m_{210}^{(k)} + 2m_{212}^{(k)}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$
$d_3$	$2m_{001}^{(k)} + 2m_{021}^{(k)} + 2m_{201}^{(k)} + 2m_{221}^{(k)}$	$m_{000}^{(k)} + m_{002}^{(k)} + m_{020}^{(k)} + m_{022}^{(k)} + m_{200}^{(k)} + m_{202}^{(k)} + m_{220}^{(k)} + m_{222}^{(k)}$

doi:10.1371/journal.pone.0011384.t005

**Table 6.** The  $\chi^2$  test statistics calculated to test the additive and dominant effects at each SNP genotyped from DLG5.

SNP	Additive	Dominant
	(a)	(d)
Arg30Gln	1.196	<b>14.316</b> (0.00015)
Glu514Gln	0	0.355
Pro979Leu	0	<b>6.095</b> (0.0136)
Gly1066Gly	0.718	<b>4.297</b> (0.0382)
Pro1371Gln	0	1.933

The *p*-values for those significant effects (in boldface) are given in parentheses. doi:10.1371/journal.pone.0011384.t006

display higher three-locus significance level than the other SNPs. Arg30Ln, Glu514Gln, and Pro979Leu produce the most numerous forms of epistasis (3), followed by the combinations of Arg30Ln, Gly1066Gly, and Pro1371Gln (2), Glu514Gln, Gly1066Gly, and Pro1371Gln (2), Arg30Ln, Glu514, and Pro1371Gln (1), Arg30Ln, Pro979Leu, and Pro1371Gln (1), Pro979Leu, Gly1066Gly, and Pro1371Gln (1). The three SNPs with significant main effects (Arg30Ln, Pro979Leu, and Pro1371Gln) do not produce a significant three-locus epistatic interaction. The two SNPs displaying non-significant main effects (Glu514Gln and Pro1371Gln) could generate significant three-locus interactions with SNPs Arg30Ln and Gly1066Gly but not with Pro979Leu (table 8).

After significant high-order epistasis is detected, the next step is to make a biological interpretation of such epistasis. To interpret it, we will use the dominant (*d*) × additive (*a*) × additive (*a*) epistasis among Arg30Ln, Glu514Gln, and Pro979Leu as an example. Table 9 gives the structure of genetic effects for each three-locus genotypic value in terms of the additive, dominant, and epistatic

**Table 7.** The  $\chi^2$  test statistics calculated to test the two-SNP epistasis between each pair of SNPs genotyped from DLG5.

SNP Pair	<i>a</i> × <i>a</i>	<i>a</i> × <i>d</i>	<i>d</i> × <i>a</i>	<i>d</i> × <i>d</i>
Arg30Gln × Glu514Gln	0.113	0.112	<b>3.292</b> (0.040)	<b>2.909</b> (0.020)
Arg30Gln × Pro979Leu	0.118	1.085	<b>3.958</b> (0.025)	<b>2.405</b> (0.040)
Arg30Gln × Gly1066Gly	0.005	1.393	1.453	1.741
Arg30Gln × Pro1371Gln	0.107	0.097	<b>3.184</b> (0.050)	<b>2.740</b> (0.027)
Glu514Gln × Pro979Leu	0	1.211	0.314	0.340
Glu514Gln × Gly1066Gly	0.222	1.160	0.545	1.205
Glu514Gln × Pro1371Gln	0	0.500	0.107	0.567
Pro979Leu × Gly1066Gly	0.261	0.920	0.001	0.607
Pro979Leu × Pro1371Gln	0	0.401	1.434	0.045
Gly1066Gly × Pro1371Gln	0.290	1.584	1.543	1.907

The *p*-values for those significant effects (in boldface) are given in parentheses. doi:10.1371/journal.pone.0011384.t007

effects of different orders. The *d* × *a* × *a* epistasis only contributes to the genotypic value of *AaBBCC*, *AaBBcc*, *AabbCC*, and *Aabbcc* (table 9). For each of these four genotypes, their values are partitioned into different effect components for both cases and controls (table 10). As can be seen, the *d* × *a* × *a* epistasis increases, by 9 cases, the incidence of IBD for those with genotype *AaBBCC* or *Aabbcc*, but decreases the IBD incidence of those carrying genotype *AaBBcc* or *AabbCC* with the same extent.

### Computer Simulation

Simulation studies were undertaken to examine the statistical behavior of the new model. We will focus on the investigation of the power and false positive rates (FDR) for the detection of three-locus epistasis. Three different simulation schemes will be used with varying numbers of cases vs. controls, 200 vs. 200, 400 vs. 400, and 1000 vs. 1000. The eight possible forms of three-locus epistasis can be sorted into four presentative ones, (1) additive × additive × additive (no dominant effect), (2) additive × additive × dominant, additive × dominant × additive, and dominant × additive × additive (no one dominant effect), (3) additive × dominant × dominant, dominant × additive × dominant, and dominant × dominant × additive (two dominant effects), and (4) dominant × dominant × dominant (three dominant effects).

For a real data set, different SNPs may be associated or independent of each other. We will investigate how SNP-SNP associations affect the behavior of the new model. In one data set, three SNPs with the same allele frequency were simulated with pair-wise and three-locus linkage disequilibria. Among the three SNPs, only additive × additive × additive, additive × additive × dominant, additive × dominant × dominant, and dominant × dominant × dominant were assumed to exist. This can be done by simulating a contingency table with constraints  $\chi^2_{a_1a_2a_3} > \chi^2_{0.05} = 3.84$ ,  $\chi^2_{a_1a_2d_3} > \chi^2_{0.05} = 3.20$ ,  $\chi^2_{a_1d_2d_3} > \chi^2_{0.05} = 3.20$ ,  $\chi^2_{d_1d_2d_3} > \chi^2_{0.05} = 2.14$  and the test statistics for the other effects > the corresponding thresholds. The same parameters, except that there is no linkage disequilibrium, were used to simulate the second data set containing three SNPs.

Table 11, table 12, and table 13 give the power and false positive error rates (FPR) of the three-locus interaction detection by the new epistatic models. The power to detect the three-locus epistasis increase remarkably with sample size in a case-control study. With sample sizes of 200 vs. 200, there is power of about 0.51–0.61, with the additive × additive × additive epistasis detected most easily, followed by the additive × additive × dominant epistasis, the additive × dominant × dominant epistasis, and the dominant × dominant × dominant epistasis. When sample sizes increase to 400 vs. 400, the power for the three-locus epistasis detection will surpass three quarters. If sample sizes 1000 vs. 1000 are used, the power reaches 0.99 or more. In general, whether the SNPs are associated or independent does not affect the power substantially, although in some cases the power is higher for associated SNPs than independent SNPs.

The power displays a small FPR (tables 11, 12, and 13). Even if small sample sizes 200 vs. 200 are used, there is still a small chance that the model provides a false positive result for the three-locus epistasis detection. The FPR was found to be consistent, regardless of sample sizes and the degree of SNP-SNP associations.

### Discussion

The phenotypic variation of a trait or disease is highly complex given its polygenic inheritance and environmental influence. Most original quantitative genetic models generally assume that allelic effects are additive, with the size linearly proportional to the

**Table 8.** The  $\chi^2$  test statistics calculated to test the three-SNP epistasis between each pair of SNPs genotyped from DLG5.

SNP Triplet	$a \times a \times a$	$a \times a \times d$	$a \times d \times a$	$a \times d \times d$	$d \times a \times a$	$d \times a \times d$	$d \times d \times a$	$d \times d \times d$
Arg30Ln × Glu514Gln × Pro979Leu	0.465	1.274	0.008	0.202	<b>7.437</b>	<b>2.780</b>	<b>3.291</b>	1.533
					(0.0025)	(0.045)	(0.025)	
Arg30Ln × Glu514Gln × Gly1066Gly	0.010	3.038	0.054	1.499	2.469	0.340	0.815	0.134
Arg30Ln × Glu514Gln × Pro1371Gln	0.426	0.390	0.027	0.183	<b>5.818</b>	2.040	2.329	1.011
					(0.008)			
Arg30Ln × Pro979Leu × Gly1066Gly	0.002	2.460	0.576	0.772	3.140	0.250	1.151	0.057
Arg30Ln × Pro979Leu × Pro1371Gln	0.448	0.315	1.601	0.014	<b>7.061</b>	2.411	2.438	1.127
					(0.0035)			
Arg30Ln × Gly1066Gly × Pro1371Gln	0.005	1.880	<b>4.250</b>	<b>2.618</b>	3.076	1.652	0.814	0.644
			(0.020)	(0.050)				
Glu514Gln × Pro979Leu × Gly1066Gly	1.101	2.096	0.046	0.908	1.447	1.158	0.120	0.774
Glu514Gln × Pro979Leu × Pro1371Gln	0	0.858	2.262	0.009	0.687	0.672	0.262	0.015
Glu514Gln × Gly1066Gly × Pro1371Gln	1.191	<b>3.457</b>	<b>3.324</b>	1.994	1.180	1.832	1.716	1.817
		(0.040)	(0.045)					
Pro979Leu × Gly1066Gly × Pro1371Gln	1.322	<b>3.298</b>	2.727	1.756	0.100	0.436	1.082	1.255
		(0.050)						

The *p*-values for those significant effects (in boldface) are given in parentheses.  
doi:10.1371/journal.pone.0011384.t008

**Table 9.** Genetic effect components of different three-locus genotypic values.

Genotype	$a_1$	$a_2$	$a_3$	$d_1$	$d_2$	$d_3$	$i_{a_1 a_2}$	$i_{a_2 a_3}$	$i_{a_1 a_3}$	$i_{d_1 d_2}$	$i_{d_2 d_3}$	$i_{d_1 d_3}$	$i_{a_1 d_2}$	$i_{a_1 d_3}$	$i_{a_2 d_3}$	$i_{d_1 a_2}$	$i_{d_1 a_3}$	$i_{d_2 a_3}$	$i_{a_1 a_2 a_3}$	$i_{d_1 a_2 a_3}$	$i_{a_1 d_2 a_3}$	$i_{a_1 a_2 d_3}$	$i_{a_1 d_2 d_3}$	$i_{d_1 a_2 d_3}$	$i_{d_1 d_2 a_3}$	$i_{d_1 d_2 d_3}$
<b>AABBCC</b>	+	+	+				+	+	+										+							
<b>AABBcC</b>	+	+					+	+					+	+												+
<b>AABbCC</b>	+	+	-				+	-	-										-							
<b>AABbCc</b>	+			+						+			+													+
<b>AABbcc</b>	+	-		+									+						-							-
<b>AAbbCC</b>	+	-	+				-	-	+										-							
<b>AAbbCc</b>	+	-			+								+	-												-
<b>AAbbcc</b>	+	-	-				-	+	-										+							
<b>AaBBCC</b>		+	+							+					+				+							
<b>AaBBcC</b>		+			+							+		+	+											+
<b>AaBBcc</b>		+	-											+					-							-
<b>AaBbCC</b>			+	+	+													+	+							+
<b>AaBbCc</b>					+	+						+	+	+												+
<b>AaBbcc</b>			-	+	+														-							-
<b>AabbCC</b>		-	+	+															-							-
<b>AabbCc</b>		-			+								+	-	-											-
<b>Aabbcc</b>		-	-	+															-							-
<b>aaBBCC</b>	-	+	+				-	+	-																	-
<b>aaBBcC</b>	-	+				+													-	+						-
<b>aaBBcc</b>	-	+	-				-	-	+											+						
<b>aaBbCC</b>	-		+		+													+								-
<b>aaBbCc</b>	-				+	+						+	-	-												-
<b>aaBbcc</b>	-		-		+														-							+
<b>aabbCC</b>	-	-	+				+	-	-											+						
<b>aabbCc</b>	-	-				+	+																			+
<b>aabbcc</b>	-	-	-				+	+	+																	

The genotypic value containing the dominant × additive × additive epistasis are in boldface.  
doi:10.1371/journal.pone.0011384.t009

**Table 10.** The genetic effect components of four particular genotypes, *AaBBCC*, *AaBBcc*, *AabbCC*, and *Aabbcc* at three SNPs, Arg30L<sub>n</sub>, Glu514Gln, and Pro979Leu, which contain the dominant × additive × additive three-locus epistasis.

Genotype		$a_2$	$a_3$	$d_1$	$i_{a_2a_3}$	$i_{d_1a_2}$	$i_{d_1a_3}$	$i_{d_1a_2a_3}$
<i>AaBBCC</i>	Cases	-16	16	-11	-16	11	-11	11
	Controls	-10	10	-2	-10	2	-2	2
	Difference	(-6)	(6)	(-9)	(-6)	(9)	(-9)	(9)
<i>AaBBcc</i>	Cases	-16	-16	-11	16	11	-11	-11
	Controls	-10	-10	-2	10	2	-2	-2
	Difference	(-6)	(-6)	(-9)	(6)	(9)	(-9)	(-9)
<i>AabbCC</i>	Cases	16	16	-11	16	-11	-11	-11
	Controls	10	10	-2	10	-2	-2	-2
	Difference	(6)	(6)	(-9)	(6)	(-9)	(-9)	(-9)
<i>Aabbcc</i>	Cases	16	-16	-11	-16	-11	-11	11
	Controls	10	-10	-2	-10	-2	-2	2
	Difference	(6)	(-6)	(-9)	(-6)	(-9)	(-9)	(9)

doi:10.1371/journal.pone.0011384.t010

number of alleles. These models are modified by considering that there are genetic interactions between different alleles at the same locus (dominance). It is now recognized that the interactions between different loci (epistasis) within gene networks may play an important role [14,15]. More recent evidence shows that high-order epistasis among more than two genes may form a crucial component in genetic interaction networks [9,11,17,18]. In fact, quantitative genetic analyses have detected high-order epistatic effects in plants. For example, high-order epistasis could be correlated with the aggressiveness of the isolate of *Phytophthora capsici* through influencing double crosses among different loci at meiosis [28]. Wu [29] used a mating design with clonal replicates to identify the significant contribution of high-order epistasis to genetic variation in stem wood growth traits in poplars.

An increasing availability of high-throughput SNP data has led to the development of various statistical approaches for effectively analyzing epistasis among multiple polymorphisms, including logistic regression, multifactor dimensionality reduction (MDR), Bayesian analysis, and machine learning [15,21,30–32]. In this article, we developed a general model for detecting the epistasis of any order in case-control genetic association studies by integrating traditional quantitative genetic principles. Despite the existence, high-order epistasis may be obscured by metabolic network redundancy [17]. The integration of quantitative genetic princi-

**Table 11.** Power and false positive rates (FPR) for the detection of three-locus epistasis among associated and independent SNPs for 200 cases and 200 controls.

Epistasis	Associated		Independent	
	Power	FPR	Power	FPR
Additive × additive × additive	61.2	4.9	51.8	4.0
Additive × additive × dominant	56.7	4.8	45.2	5.3
Additive × dominant × dominant	49.3	5.1	48.4	4.7
Dominant × dominant × dominant	51.0	6.0	56.1	6.8

doi:10.1371/journal.pone.0011384.t011

**Table 12.** Power and false positive rates (FPR) for the detection of three-locus epistasis among associated and independent SNPs for 400 cases and 400 controls.

Epistasis	Associated		Independent	
	Power	FPR	Power	FPR
Additive × additive × additive	85.0	5.8	79.6	5.4
Additive × additive × dominant	84.2	4.5	78.6	6.0
Additive × dominant × dominant	77.6	5.6	76.8	7.5
Dominant × dominant × dominant	79.8	7.5	86.6	6.0

doi:10.1371/journal.pone.0011384.t012

ples makes our approach capable to identify high-order epistatic interactions with genetic relevance. The model was tested by simulation studies. It displays adequate power for the detection of high-order epistasis with a modest sample size; for example, 400 cases vs. 400 controls. When sample sizes of cases and controls increase to 1000 vs. 1000, which is currently not a problem for most genetic association studies, the model has almost full power to detect three-locus epistasis of different forms. Even if a small size of samples (say 200 vs. 200), the new model has a low false positive rate for epistatic detection. The practical application of the model is validated by analyzing a real data set for the genetic study of inflammatory bowel disease. The model detected significant three-locus epistatic interactions among different SNPs genotyped from a candidate gene DLG5 [27].

Our model allows the characterization of epistasis of any order. Its implementation into a practical setting of genome-wide association studies is challenged by an exponentially increasing number of SNP-SNP combinations. To make this tractable, one may incorporate optimization techniques into our model, allowing the selection of the most important combinations. An additional issue is to determine the critical threshold with multiple correlated SNPs in genome-wide association studies. An empirical approach for determining a genome-wide threshold is to employ non-parametric permutation testing (see ref. [21,30,33–35]). Lastly, the model is developed to detect multilocus epistasis at the SNP level, but given recent discoveries for the importance of haplotypes in trait control [36–39], the model should be extended to consider high-order interactions expressed by different haplotypes. In the current model specification, we choose controls that are matched for cases in terms of biological, environmental, or demographical factors. When such matches are not possible, we need to embed these factors as covariates into the model, in which the interactions between genes and these factors can be tested. Third, the model

**Table 13.** Power and false positive rates (FPR) for the detection of three-locus epistasis among associated and independent SNPs for 1000 cases and 1000 controls.

Epistasis	Associated		Independent	
	Power	FPR	Power	FPR
Additive × additive × additive	99.8	5.9	99.1	5.6
Additive × additive × dominant	99.6	4.1	99.3	6.3
Additive × dominant × dominant	99.2	5.8	98.0	5.9
Dominant × dominant × dominant	98.6	8.7	99.7	6.6

doi:10.1371/journal.pone.0011384.t013



can be extended with multiple diseases to consider the pleiotropic effect of a gene. The results about high-order epistasis detection using the this and extended models could be used for iterative model building and functional annotation of genes. Future applications of these results includes analysis of the metabolic networks of pathogenic organisms and generation of epistatic candidate models for genome-wide association studies.

## Supporting Information

**Text S1** Mathematical algorithm.

## References

- Weng G, Bhalla U, Iyengar R (1999) Complexity in biological signaling systems. *Science* 284: 92–96.
- Hlavacek W, Faeder J (2009) The Complexity of Cell Signaling and the Need for a New Mechanics. *Science's STKE* 2.
- Huang L, Sternberg P (1995) Genetic dissection of developmental pathways. *Methods in cell biology* 48: 97–122.
- McMullen M, Byrne P, Snook M, Wiseman B, Lee E, et al. (1998) Quantitative trait loci and metabolic pathways. *Proceedings of the National Academy of Sciences of the United States of America* 95: 1996.
- Ritchie M, Hahn L, Roodi N, Bailey L, Dupont W, et al. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *The American Journal of Human Genetics* 69: 138–147.
- Martin M, Gao X, Lee J, Nelson G, Detels R, et al. (2002) Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature genetics* 31: 429–434.
- Gabutero E, Moore C, Mallal S, Stewart G, Williamson P (2007) Interaction between allelic variation in IL12B and CCR5 affects the development of AIDS: IL12B/CCR5 interaction and HIV/AIDS. *AIDS* 21: 65.
- Hirschhorn J, Daly M (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6: 95–108.
- Marchini J, Donnelly P, Cardon L (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature genetics* 37: 413–417.
- Wang W, Barratt B, Clayton D, Todd J (2005) Genome-wide association studies: theoretical and practical concerns. *Nature Reviews Genetics* 6: 109–118.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81: 559–575.
- Wan X, Yang C, Yang Q, Xue H, Tang N, et al. (2010) Predictive rule inference for epistatic interaction detection in genome-wide association studies. *Bioinformatics* 26: 30.
- Phillips P (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* 9: 855–867.
- Moore J, Williams S (2009) Epistasis and its implications for personal genetics. *The American Journal of Human Genetics* 85: 309–320.
- Stich B, Yu J, Melchinger A, Piepho H, Utz H, et al. (2007) Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. *Genetics* 176: 563.
- Beerenwinkel N, Pachter L, Sturmfels B, Elena S, Lenski R (2007) Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evolutionary Biology* 7: 60.
- Imielinski M, Belta C (2008) Exploiting the pathway structure of metabolism to reveal high-order epistasis. *BMC Systems Biology* 2: 40.
- Hansen T, Wagner G (2001) Epistasis and the mutation load: a measurement-theoretical approach. *Genetics* 158: 477.
- Zhang Y, Liu J (2007) Bayesian inference of epistatic interactions in case-control studies. *Nature genetics* 39: 1167–1173.
- Nunkesser R, Bernholt T, Schwender H, Ickstadt K, Wegener I (2007) Detecting high-order interactions of single nucleotide polymorphisms using genetic programming. *Bioinformatics* 23: 3280.
- Gayán J, González-Pérez A, Bermudo F, Sáez M, Royo J, et al. (2008) A method for detecting epistasis in genome-wide studies using case-control multi-locus association analysis. *BMC genomics* 9: 360.
- Bateson W (1910) Mendels principles of heredity. *Molecular and General Genetics MGG* 3: 108–109.
- Kempthorne O (1968) The correlation between relatives on the supposition of mendelian inheritance. *American Journal of Human Genetics* 20: 402.
- Workman P (1973) Biometrical genetics. The study of continuous variation. *American Journal of Human Genetics* 25: 461.
- Liu T, Thalamuthu A, Liu J, Chen C, Wu R (2010) A model for testing epistatic interactions of complex diseases in case-control studies. *Biostatistics* (in press).
- Lin Z, Poritz L, Franke A, Li T, Ruether A, et al. (2009) Genetic association of DLG5 R30Q with familial and sporadic inflammatory bowel disease in men. *Disease markers* 27: 193–201.
- Stoll M, Corneliusen B, Costello C, Waetzig G, Mellgard B, et al. (2004) Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nature genetics* 36: 476–480.
- Bartual R, Lacasa A, Marsal J, Tello J (1993) Epistasis in the resistance of pepper to phytophthora stem blight (*Phytophthora capsici* L.) and its significance in the prediction of double cross performances. *Euphytica* 72: 149–152.
- Wu R (1996) Detecting epistatic genetic variance with a clonally replicated design: models for low/high-order nonallelic interaction. *TAG Theoretical and Applied Genetics* 93: 102–109.
- Liang Y, Kelemen A (2008) Statistical advances and challenges for analyzing correlated high dimensional SNP data in genomic study for complex diseases. *Statistics Surveys* 2: 43–60.
- Kayano M, Takigawa I, Shiga M, Tsuda K, Mamitsuka H (2009) Efficiently finding genome-wide three-way gene interactions from transcript- and genotype-data. *Bioinformatics* 25: 2735.
- Jiang R, Tang W, Wu X, Fu W (2009) A random forest approach to the detection of epistatic interactions in case-control studies. *BMC bioinformatics* 10: S65.
- Carlborg O, Andersson L (2002) Use of randomization testing to detect multiple epistatic QTLs. *Genetics Research* 79: 175–184.
- Alison M (2009) The effect of alternative permutation testing strategies on the performance of multifactor dimensionality reduction. *BMC Research Notes* 1.
- Edwards T, Turner S, Torstenson E, Dudek S, Martin E, et al. (2010) A General Framework for Formal Tests of Interaction after Exhaustive Search Methods with Applications to MDR and MDR-PDT.
- Judson R, Stephens J, Windemuth A (2000) The predictive power of haplotypes in clinical response. *pgs* 1: 15–26.
- Bader J (2001) The relative power of SNPs and haplotype as genetic markers for association tests. *pgs* 2: 11–24.
- Liu T, Johnson J, Casella G, Wu R (2004) Sequencing complex diseases with HapMap. *Genetics* 168: 503.
- Rha S, Jeung H, Choi Y, Yang W, Yoo J, et al. (2007) An association between RRM1 haplotype and gemcitabine-induced neutropenia in breast cancer patients. *The Oncologist* 12: 622.

Found at: doi:10.1371/journal.pone.0011384.s001 (0.14 MB PDF)

## Author Contributions

Conceived and designed the experiments: RW. Performed the experiments: ZL JPH. Analyzed the data: ZW TL. Contributed reagents/materials/analysis tools: ZW TL ZL JPH. Wrote the paper: RW.