

NIH Public Access

Author Manuscript

Genet Epidemiol. Author manuscript; available in PMC 2013 June 25

Published in final edited form as:

Genet Epidemiol. 2009; 33(01): S58–S67. doi:10.1002/gepi.20474.

The Challenge of Detecting Epistasis (G×G Interactions): Genetic Analysis Workshop 16

Ping An¹, Odity Mukherjee², Pritam Chanda³, Li Yao⁴, Corinne D Engelman⁴, Chien-Hsun Huang⁵, Tian Zheng⁵, Ilija P Kovac⁶, Marie-Pierre Dubé⁶, Xueying Liang⁷, Jia Li⁸, Mariza de Andrade⁸, Robert Culverhouse⁹, Doerthe Malzahn¹⁰, Alisa K Manning¹¹, Geraldine M Clarke¹², Jeesun Jung¹³, and Michael A Province¹

¹ Division of Statistical Genomics and Department of Genetics, Washington University School of Medicine, St. Louis, MO

² National Center for Biological Sciences, Bangalore, India

³ Department of Computer Science and Engineering, State University of New York, Buffalo, NY

⁴ Department of Human Development and Family Studies, University of Wisconsin-Madison, Madison, WI

- ⁵ Department of Statistics, Columbia University, New York, NY
- ⁶ Research Center of the Montreal Heart Institute, Montreal, Quebec, Canada

⁷ Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD

- ⁸ Department of Health Sciences Research, Mayo Clinic, Rochester, MN
- ⁹ Department of Medicine, Washington University School of Medicine, St. Louis, MO

¹⁰ Department of Genetic Epidemiology, University Medical Center, University of Goettingen, Germany

- ¹¹ School of Public Health, Boston University, Boston, MA
- ¹² Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

¹³ Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN

Abstract

Interest is increasing in epistasis as a possible source of the unexplained variance missed by genome-wide association studies. The Genetic Analysis Workshop 16 Group 9 participants evaluated a wide variety of classical and novel analytical methods for detecting epistasis, in both the statistical and machine learning paradigms, applied to both real and simulated data. Because the magnitude of epistasis is clearly relative to scale of penetrance, and therefore to some extent, to the choice of model framework, it is not surprising that strong interactions under one model might be minimized or even disappear entirely under a different modeling framework.

Corresponding Author: Michael A Province, Division of Statistical Genomics, Box 8506, 4444 Forest Park Blvd, Washington University School of Medicine, St. Louis, Missouri 63108 USA, mprovince@wustl.edu.

Keywords

generalized linear model; machine learning methods

INTRODUCTION

The term "epistasis" was first described by the English geneticist William Bateson [1907] to denote the suppression of gene expression at one locus by a gene at another locus. However, modern geneticists more often think of epistasis using Fisher's [1918] conceptualization, as a departure from additivity in the penetrance for two or more loci, in the same way that dominance is a departure from additivity in the penetrance at one locus. Thus, penetrance models of epistasis require additional interaction terms (each with its own corresponding parameter) in and above the additive "main effect" terms for each locus. In this framework a test of epistasis is a test of whether these gene-gene ($G \times G$) interaction term(s) are zero or not, and lack of epistasis represents a special class of all possible multi-locus penetrance functions. It is of course an empirical question whether epistasis plays a major or minor role for any given trait in any particular population or defined subsample. But interest has lately been increasing in epistasis as one possible source of the so called "dark matter" or "missing R^{2*} in genome-wide association scans (GWAS) for complex traits (i.e., the fact that the cumulative main effects from GWAS signals account for far less of the total predictive R^2 than the estimated heritabilities of these traits). Correspondingly, analytical methods to detect and estimate the degree of epistasis are becoming more sophisticated and more numerous. In an attempt to compare and contrast the advantages and disadvantages of various epistatic models and methods of detection, investigators participating in Genetic Analysis Workshop 16 (GAW16) Group 9 applied a number of traditional as well as novel methods to three large, complex trait data sets.

Data

Eight groups analyzed the North American Rheumatoid Arthritis Consortium rheumatoid arthritis (RA) data (GAW16 Problem 1 [Amos et al., 2009]): Chanda et al. [2009], Clarke et al. [2009], Huang et al. [2009], Jung et al. [2009], Li et al. [2009], Liang et al. [2009], Manning et al. [2009], and Mukherjee et al. [2009]. Three groups analyzed the Framingham Heart Study (FHS) data (GAW16 Problem 2 [Cupples et al., 2009]): An et al. [2009], Malzahn et al. [2009], and Yao et al. [2009]. Finally, three groups analyzed the simulated data which was based upon FHS (GAW16 Problem 3 [Kraja et al., 2009]): Culverhouse et al. [2009], Kovac and Dube [2009], and Malzahn et al. [2009]. While most groups assessed discrete traits (RA, coronary heart disease, type 2 diabetes), five groups also focused on quantitative traits including triglyceride/high-density lipoprotein ratio [An et al., 2009], anticyclic citrinullated peptide and rheumatoid factor IgM [Mukherjee et al., 2009], low-density lipoprotein [Kovac and Dube, 2009], coronary artery calcification (CAC) and coronary event [Culverhouse et al., 2009], and CAC and body mass index [Malzahn et al., 2009]. Two groups assessed longitudinal quantitative traits in the FHS data including changes in triglyceride/high-density lipoprotein ratio [An et al., 2009] and changes in CAC and body mass index [Malzahn et al., 2009].

Methods for detecting epistasis

Approaches to detecting epistasis can be classified as either **statistical** or **machine learning** methods. Statistical methods make formal models of stochasticity or randomness, and most propose formal hypothesis tests of epistasis. By contrast, machine learning methods [Alpaydin, 2004] tend to be more heuristic, data-mining techniques that do not necessarily rely on formal statistical tests, but concentrate on efficient algorithms to identify epistatic

patterns in high-dimensional spaces, such as the space of all possible $G \times G$ interactions among a set of candidate genes, for instance. Some machine learning methods do build their search algorithm around formal statistical models. An early example of such a technique is stepwise regression. Each stage in the "variable selection" model-building process is a formal statistical multiple regression model, but using the stepwise algorithm itself to add and delete new predictors is a heuristic way to reasonably search the space for all possible models, and we typically do not worry about such formal statistical issues as corrections for multiple comparisons between submodels, the possibility of partial null hypotheses, or other probability issues concerning multiple models considered simultaneously. In GAW16 Group 9, both statistical models and machine learning approaches were used to attempt to identify epistasis, as summarized in Table I, and discussed below.

Statistical methods for epistasis

The general approach of statistical epistasis methods is to take the null hypothesis as "no epistasis" (additivity), which is preferred unless there is overwhelming evidence in the data in favor of epistasis. A significance level is provided that quantifies the probability of observing the data (or more extreme data) if the null is true (i.e., no epistasis). Thus, in the typical hypothesis testing paradigm, statistical methods tend to conclude in favor of the null of no epistasis, unless there is strong evidence in favor of epistasis. This is also in keeping with standard epidemiological practice about interactions in general, in which we tend to favor simpler models with no interaction over more complex models requiring interaction, under the principle of Occam's Razor. If we can reasonably model the data assuming additivity, we prefer this over requiring epistasis to explain the data. In fact, many statistical tests for epistasis are done in the context of specified multivariate penetrance models, which tend to be one form or another of generalized linear models (GLMs).

Generalized linear model

GLM notation developed by Nelder and Wedderburn [1972] is a general framework for describing multivariable linear models. Given a stochastic random variable phenotype, Y, and a set of (fixed, non-random) predictors, X (which includes genes G_i and G_j), we assume that (Y|X) follows a specified parametric probability distribution, P, such that:

$$\mathbf{E}[Y|\mathbf{X}] = \mathbf{l}^{-1}(\mathbf{D}[\mathbf{X}]\boldsymbol{\beta}),$$

where $\mathbf{D}[\mathbf{X}]$ denotes the design matrix for \mathbf{X} , which includes as a submatrix $[G_i || G_j || G_j || G_j XG_j]$, i.e., the design matrix contains main effect terms for each of G_i and G_j (additive as well as dominance terms for each gene) as well as interaction terms between G_i and G_j (in general, four terms for additive×additive, additive×dominance, dominance×additive, and dominance×dominance). $\boldsymbol{\beta}$ is an unknown parameter vector of intercept and slopes that we estimate from the data, and we are typically interested in the β -coefficients corresponding to the epistasis terms in the model. I (.) is a known, specified link function that is 1-1 and invertible. This framework is quite general, and many of the classical models of statistics can be formulated this way by an appropriate choice of probability distribution and link function, as shown in Table II. In fact, most of the epistasis statistical models used in GAW16 can be cast into this canonical GLM formulation, which allows us to compare and contrast models.

Case-only

Clarke et al. [2009] considered modeling a binary trait as being influenced by two bi-allelic disease susceptibility loci, *F* and *G*, according to a model of joint locus effects. Here, *F*

denotes a candidate gene single-nucleotide polymorphism (SNP) and *G* denotes an "equilibrium SNP" (i.e., tag SNPs covering a region which themselves are pairwise in low linkage disequilibrium (LD) r^2 <0.2). They tested for G×G interaction between gene and equilibrium SNPs using GLM tests based on logistic odds, proportional odds, and multinomial link functions. For each model, there are two regressions: first, *F* is modeled as the outcome variable and *G* the predictor, then vice versa. The outcome variable is categorized appropriately according to the relevant model: a binary categorization for the logistic model, an ordinal categorization for the proportional odds model, and a nominal categorization for the multinomial model, which result in three different link functions in the GLM formulation. The predictor variable is categorized as an ordinal variable in all three regressions.

Family mixed model

Kovac et al. [2009] and An et al. [2009] used a family-mixed model [Borecki and Province, 2008], which is an extension of the multiple regression model, to deal with association in family data. It can overcome the problem of non-independence of residuals within pedigrees that produces inflation of type I error if one applies regular regression and ignores family relationships. This GLM uses a gaussian probability distribution and an identity link function, just as in linear regression, but includes an additional random effect component predictor for pedigrees.

Allelic scoring method

The underlying principle of this method of Jung et al. [2009] is to identify the association of allelic combination between two unlinked markers with a disease trait so that subjects are assigned an allelic score given their observed genotype information. The score is a conditional probability of obtaining the particular allelic combination given the observed genotypes at the two loci of each subject. A linear trend of proportion of cases over total number of subjects at each allelic combination can be modeled using an extension of the Cochran-Armitage trend regression.

Omnibus test (OT)

Liang et al. [2009] applied the OT of Chatterjee et al. [2006] to detect epistasis. The omnibus method tests for gene-based effects by considering all SNPs in a given gene or region as a single group and evaluates this gene assuming a second known gene or other risk factor plays a role. Specifically, the method forms L(G) latent factors from linear combinations of G loci, and tests the GLM $\mathbf{E}[\mathbf{Y}|\mathbf{L}(\mathbf{G})] = \mathbf{1}^{-1} (\mathbf{D}[\mathbf{L}(\mathbf{G})] \boldsymbol{\beta})$ with $L_m \times L_n$ latent interactions. It then infers $G_i \times G_j$ interactions from $L_m \times L_n$ interactions and latent path loadings. The application to GAW16 Problem 1 used a logistic regression approach (binomial distribution with a logit link in the GLM) but the significance of the test gene effect includes both the main effect and the interaction between this gene and the known risk factor or gene. For the genes identified by these methods, logistic regression was used to test formally whether the interaction terms were significant predictors.

Principal-component analysis (PCA)

Li et al. [2009] extended the original PC approach to test for association between disease and multiple SNPs in a candidate gene in order to incorporate a test for G×G interaction. The procedure involves the following steps. 1) Let g_{lk} be the number of minor alleles at SNP *k* for *I*^h subject, I = 1, ..., N, k = 1, ..., K. 2) Calculate the correlation matrix *R*, where $R_{ij} = cor(g_i, g_j)$ and g_i and g_j represent the genotypes of all subjects for SNP *i* and SNP *j*, respectively. 3) Decompose *R* by singular value decomposition: $R = A \wedge A^T$. 4) Determine the factor loading by $L = A \sqrt{\Lambda}$. 5) Determine the PCs by PC = *GA*, where *G* is the standardized $N \times K$ matrix of genotypes. The standardized genotypes are calculated as:

$$\frac{g_{lk} - \overline{g}_k}{sd(g_k)}, \text{ where } \overline{g}_k = \frac{\sum_{l=1}^N g_{lk}}{N} \text{ is the mean genotype across subjects and} \\ sd(g_k) = \sqrt{\sum_{l=1}^N (g_{lk} - \overline{g}_k)^2 / (N-1)} \text{ is the standard deviation.}$$

Li et al. [2009] used PCs that explained at least 80% of the variation as the gene representation to perform a G×G interaction analysis by applying logistic regression to test for interaction between every combination of two PCs. Once significant PC interactions were identified, PC loadings were used to determine the influence of a specific SNP on the PCs because the loading represents the correlation of a SNP with a component.

Interactions among main effect genes and interactions among pathway genes

In a GWAS, or even when there are a large number of candidate genes, there can be too many possible $G \times G$ interactions to evaluate exhaustively. Manning et al. [2009] utilized two complementary approaches to reduce the number of possible $G \times G$ interactions to test. The first strategy is to use a two-stage approach test for interactions only among genes that show significant main effects. In Stage 1, a set of additive GLMs are fit, one variant at a time, and the subset of variants that show significance are selected for further consideration. In Stage 2, a series of two-variant GLMs are refit, which include every pairwise combination of the main effect subset as well as their respective interactions. The second approach, interactions among pathway genes, is similar in spirit and design. Again, a subset of genes is selected in Stage 1 and interactions are only evaluated among genes in that subset, but in this strategy, the Stage 1 subset is selected based on external biological knowledge that genes belong to the same relevant pathway, rather than based internal statistical tests from the data itself, interactions among pathway genes.

Two-step approach

Li et al. [2009] modified the two-step approach of Murcray et al. [2009] to detect geneenvironment interaction to be applied to detect $G \times G$ interaction. In the first step, a GLM model is fit predicting *G* from *F* in the combined case-control data, using the approximate method to detect epistasis in PLINK (note that this analysis does not involve phenotype, only the two genotypes). This can be considered as a modified version of the case-only approach for epistasis. Only those SNPs that show significant epistasis in Step 1 are carried forward to Step 2, in which a saturated logistic model (GLM with binomial distribution and logit link) is fit. The test of the G×G product term is the final test of epistasis.

Longitudinal nonparametric association test (LNPT)

Malzahn et al. [2009] adopted a test statistic from the area of clinical studies [Brunner et al., 2001]. The LNPT tests for association of longitudinal quantitative traits with respect to a set of influencing factors. The latter divide the cohort into subgroups. The LNPT tests the null hypothesis of no difference in trait distribution **F** between these subgroups H_0^F : **CF=0**, where **C** is a contrast matrix and $\mathbf{F} = \{F_t^{kls}\}$ is the set of distribution functions F_t^{kls} ordered by observational time point *t* and the influencing factors (*kls*) of interest (for example, two factors *k*, I = 0, 1, 2 for SNP genotype at two bi-allelic loci and additional factors for covariates, e.g., sex $s = \{m, w\}$). The LNPT test statistic is invariant with respect to monotone trait transformations. The LNPT is not a GLM because no distributional assumptions are made about **F** and the test is not restricted to contrasts of expected values. However, its test statistic resembles a heteroscedastic repeated measures GLM ANOVA, which is performed

on the mid-ranks of the longitudinal traits, estimating longitudinal covariance from the ranks without assuming any structure. The LNPT requires that individuals be followed up at the same time intervals, but individuals with partially missing values for the longitudinal phenotype can be included for computation of the test statistic. The LNPT yields a set of adjusted *p*-values for tests of average effects of the loci, covariates (e.g., sex), number of exam, and for tests of all interactions.

Cox model

Malzahn et al. [2009] converted longitudinal quantitative traits to event-time data, testing for evidence of G×G interactions with the established semi-parametric Cox model of survival analysis [e.g., Therneau and Grambsch, 2000]. Event time was defined as age at the first exam when the longitudinal trait crossed a predefined threshold. Event times are invariant with respect to monotone transformations of the trait. The Cox model estimates hazard ratios HR(G_i) to quantify the impact of a genotype G_i on event-time. Significance of G×G interactions (rejection of H₀: HR(G_1 , G_2)=HR(G_1)HR(G_2)) was evaluated with the likelihood ratio test.

UNPHASED

Kovac et al. [2009] utilized a haplotype approach to epistasis, as implemented in the program UNPHASED [Dudbridge, 2008]. It uses a likelihood framework for primarily haplotype-based analysis of data, which can include both familial and unrelated subjects. The test for $G \times G$ interaction for a quantitative trait compares the null hypothesis of equal contributions for all gene combinations (in haplotype form) sharing the same alleles at the conditioning marker, versus the alternative hypothesis of differential multiplicative contributions from the test marker. The test uses a likelihood-ratio chi-square statistics to compare models with and without the interaction terms.

Machine learning methods for epistasis

Complementary to the statistical methods for epistasis are the machine learning ones, which typically are high-dimensional heuristic search algorithms to detect G×G interactions that mostly rely upon split samples with cross validation to avoid fitting to noise. Some use the basic GLM to evaluate particular interactions, but many of these detection methods do not utilize a formal G×G model per se, and therefore do not provide an estimate of effect size. The emphasis is on efficient search among a large number of possible G×G interactions to determine which are signals and which are noise, rather than on detailed modeling of any particular G×G interaction.

Multifactor dimensionality reduction (MDR) and generalized MDR (GMDR)

Mukherjee et al. [2009] applied the GMDR method, a score-based extension of the MDR [Ritchie et al., 2001]. In MDR, multi-locus genotype combinations are classified as high-risk or low-risk genotype combinations using a threshold on the ratios of cases to controls in each combination. The best model is selected as the combination of marker with maximum cross-validation consistency and minimum prediction error. GMDR generalizes this framework, replacing the ratio of cases to controls by scores in each cell to discriminate between high risk and low risk. The GMDR algorithm allows for increased flexibility to use covariates, handling both dichotomous and continuous phenotypes, and a variety of population-based study designs such as using unbalanced case control samples.

Restricted partitioning method (RPM)

Culverhouse et al. [2009] applied the RPM [Culverhouse et al., 2004], which reduces the dimensionality of genotype comparisons by using a multiple comparisons ANOVA to

evaluate whether the phenotypes associated with each (multilocus) genotype in a particular model (e.g., a two-SNP model consisting of nine genotypes) come from the same distribution. If the answer is no, the method proposes a partition of the genotypes. The test statistic is the proportion of the trait variation explained by the partition. Statistical significance is determined by permutation testing. The RPM was developed as a method for analyzing data sets consisting of unrelated subjects, and hence can be considered only an approximately correct screening tool when applied to pedigree data, such as the FHS.

Generalized, unbiased interaction detection and estimation (GUIDE)

Yao et al. [2009] utilized GUIDE [Loh, 2002], a tree building software package. GUIDE develops classification trees using three steps: 1) perform a χ^2 test to select the most significant variable to split a node; 2) select the split threshold that minimizes the node impurity measure; 3) recursively repeat Steps 1 and 2 until there are too few observations in each node. After building a complete tree, three methods are used to decide how much of the tree to retain: cross-validation pruning, test-sample pruning, and no pruning, where the criterion for judging the correct amount of pruning is that which minimizes the unbiased estimate of misclassification cost. GUIDE allows fast computation, provides a natural extension to data sets with categorical variables, and direct detection of local two-variable interactions. It has four useful properties: 1) negligible selection bias; 2) sensitivity to curvature and local pairwise interactions between regressor variables; 3) inclusion of categorical predictor variables; and 4) choice of three roles for each ordered predictor variable: split selection only, regression modeling only, or both. GUIDE can process a large number of SNPs in one run. However, it is still not feasible to run the entire Problem 2 FHS data set at one time due to computation limitations (350k SNPs after filtering for lack of Hardy-Weinberg equilibrium (p < 0.001) and low minor allele frequency < 5%, which results in a file over 10 GB in size).

Random forests (RF)

Liang et al. [2009] utilized the RF approach of Breiman [2001], which involves tree models fit to bootstrapped samples of subjects and predictors. Each bootstrap tree provides a classification, and these are aggregated as votes to form a final classification. RFs are less likely to fit to noise than are simple trees.

K-Way interaction information (KWII)

Chanda et al. [2009] developed an entropy-based method for detecting epistasis called KWII. KWII is defined as the amount of information (redundancy or synergy) present in the set of variables that is not present in any subset of these variables. Formally, if S is a set of variables that includes both predictors (including genes) as well as the response phenotype, then KWII $(S) = \sum_{t \in S} (-1)^{|S| - |t|} H(t)$, where $H(t) = \sum_{t \in u} P(t = u) \log_2 [P(t = u)]$ is the Shannon entropy information (which measures the amount of information in a system one is missing if that variable is unknown or not used). Sets of predictors with larger KWII indicate stronger interactions with the phenotype, and thus higher association. Given the strong LD patterns in the genome, there can be multiple sets of SNPs that are formally disjoint, but essentially contain the same information, which is characterized by a redundancy metric: specifically, the redundancy between sets of predictors S_1 and S_2 is given by the maximized average of pairwise LD (R^2) between variables from S_1 and S_2 red (S_1S_2) = max($\sum_{x \in S_1, y \in S_2}$ $r^2(x, y)/m$). The KWII algorithm performs a stepwise search for the set of predictors that when added to the existing set, S, produces the greatest KWII and is not redundant to S. The computational burden on checking redundancy is sufficiently high that only a fixed number of iterations can be performed, which yields a fixed number of single- and two-variable combinations and their corresponding KWII values, which are input to a second evaluation

stage. In this stage, a traditional logistic regression fully saturated GLM is performed, using a binomial distribution and a logit link.

Genotype trait distortion (GTD)

The GTD method of Huang et al. [2009] is a variation on the backward haplotype transmission association [Lo and Zheng, 2002] and backward genotype-trait association methods of Zheng et al. [2006]. Given *k* SNP markers, there are 3*k* possible unphased genotypes. The GTD statistic, *V*, is defined on the sum of squared difference between genotypes' relative frequency among the cases and controls and measures the joint effects of these *k* SNPs on the disease status. Specifically $V = \sum_{g \in G} (E[Y=1|g] - E[Y=0|g])^2$. Based on *V* as the measure of joint effect for a set of SNPs, Huang et al. [2009] measures the evidence of interaction by using the relative excess effect from a pair of SNPs over their individual marginal effects. Significance of the interaction measure is assessed by permutation testing.

RESULTS

In Figure 1, we show the epistasis detected by various methods for the GAW16 Problem 1 (RA data set), which was the data that was analyzed by most groups for epistasis. The most striking observation is the lack of consistency of results. There are a few genes that show consistent epistasis by multiple methods, such as TRAF1-C5 and PTPN22 (three methods: MDR, RF, and OT); HLA-DRB1 and PTPN22 (two methods: OT and RF), HLA-DRB1 and TRAF1-C5 (two methods: RF and OT); and HLA-B and HLA-C (two methods: GTD and KWII). Some methods found many G×G interactions that few (if any) other methods found. For instance, GMDR found 18 interactions, not one of which any other method found. GTD found 17 interactions, only one of which one other method found. Some methods found only a small number of interactions. For example, MDR identified two: HLA-C and PTPN22 as well as PTPN22 and TRAF1-C5. Either some of these methods are homing in on information not being used by other methods, or some are more powerful than others, or some are more prone to fitting to noise than others. It is difficult to reach a definitive conclusion because this is a real data example, and we therefore do not know the truth. Hence, we do not know when to praise or scold a method for either finding or not finding what it "should have." But part of the difficulty in comparing methods may arise from the relativity of epistasis to scale of penetrance.

DISCUSSION

The relativity of epistasis

Because epistasis is simply a departure from additivity in multi-locus penetrance, it has been known for some time that such statistical interactions are scale dependent [Greenland et al., 2008]. Recently, several authors [Cordell, 2002; Frankel and Schork, 1996; Greenland and Rothman, 1998] have emphasized that the choice of how one models epistasis and in particular, the scale upon which penetrance is measured, will greatly affect whether additivity is maintained and therefore whether there "is" or "is not" epistasis. In particular, by simply rescaling the problem we can "create" or "remove" epistasis. This is illustrated in Figure 2, for two hypothetical two-locus examples. In Figures 2a and 2b, we show the interaction between two genes *G* and *F* in which the probability of disease (penetrance) for the baseline genotype group is *P*[disease]*G*=*aa*, *F*=*bb*]=0.001, and each *A* allele dose for gene *G* increases the conditional probability risk by six-fold, while each *B* allele for gene *F* increases the conditional probability risk by five-fold. In Figure 2a, both genes have no dominance and no interaction when modeled on the multiplicative scale (i.e., all three *F* genotype lines when plotted against the *G* genotype on the x-axis and the log probability of

disease given genotype along the y-axis, are linear and parallel). This would utilize a $\log(P)$ link function in the GLM, which corresponds to a multiplicative risk model. However, these same data are shown in Figure 2b on the $\log(odds)$ scale as is relevant when analyzing by logistic regression (using a logit link = $\log(P'(1-P))$ in the GLM). Here, both genes show strong dominance (non-linear response to the *G* genotype by the *F* genotype) as well as strong G×G interaction (non-parallel lines). We have not changed the data, just the scale of the y-axis, and we have created epistasis. By contrast, in Figures 2c and 2d, consider two other genes *J* and *K* in which the baseline genotype group penetrance is *P*[disease | *J=cc*, *K=dd*]=0.5, (which corresponds to odds(*P*)=1), and each *C* allele dose for gene *J* increases the odds four-fold, while each *D* allele dose for gene *K* increases the odds two-fold. Both genes show strong dominance as well as G×G interaction on the $\log(P)$ scale (Figure 2c), which would be the conclusion according to a multiplicative model, but these same data show no dominance and no interaction on the $\log(odds)$ scale for logistic regression (Figure 2d). By simply rescaling the y-axis, we have removed epistasis.

Hence, for most methods, the existence of epistasis and/or dominance is dependent upon the scale of the response and therefore also on the choice of model or link function. It would therefore not be surprising that some models might find strong epistasis, while others applied to the same data might find little or no epistasis, just as we observed in GAW16. It is important to realize that this is a deeper issue than just that the "power to detect epistasis" differs by method. In Figures 2a and 2d, there is metaphysically no epistasis (not just close to zero epistasis, which might possibly be detected by some more powerful methods). We have found a scale in which it is impossible to detect epistasis, because by simply changing scale, we have flipped from the alternative to the null hypothesis (where issues of power are moot). Evidently, the term "epistasis" in the sense of Fisher is non-additivity, not some objective biological condition that exists in and of itself, outside of the way in which we must specify a particular penetrance scale, much like in physics where the ideas of "motion" and "rest" only make sense with respect to a particular frame of reference.

The reality of epistasis

Just because the concept of epistasis requires a scale or frame of reference to makes sense, it does not mean that it is an imaginary or unimportant phenomenon. No one would suggest that the concept of motion is illusory just because it is relative, nor that all things are really standing still. In fact, much of the success of classical Newtonian physics centers around embracing the frame of reference concepts in order to form strong and accurate models of bodies in motion. Indeed, the relativity of epistasis means that essentially, for any pair of genes there is at most one frame of reference, one scale, upon which additivity holds, and on all other scales there is non-additivity or epistasis, just as in our examples in Figure 2. Moreover, it is **not** true that there is always some monotonic rescaling of the penetrance function that will reduce epistasis to zero. Whenever different genotypes at one locus cause the order of the penetrances by genotype at the other locus to reverse, there can be no monotonic transformation that will "remove" epistasis, as illustrated in Figure 3. Here we reproduce two real examples of epistasis from model organisms. Carlborg et al. [2004] found this type of persistent epistasis in chicken growth (their Figure 3), and Learny et al. [2005] found it in mice for both molar size and shape (their Figure 2). This pattern of epistasis will persist, regardless of how we monotonically transform the penetrance function, and we can never find a scale on which the two genes act additively. Therefore, if anything, it might be more rightly emphasized that every pair of genes will show some degree of epistasis on almost every scale of reference (all but save at most one scale) and therefore, we should be cautious about making untested assumptions that there is no epistasis on the particular scale on which we model our data.

Acknowledgments

The Genetic Analysis Workshops are supported by NIH grant R01 GM031575 from the National Institute of General Medical Sciences. The authors acknowledge the investigators that contributed the phenotype, genotype, and simulated data. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (N01 HC25195). Creation of the simulated Framingham Heart Study data was supported by the Washington University Institute of Clinical and Translational Sciences, NIH grant 1U54RR023496. The GAW16 Framingham Heart Study and simulated data were obtained through dbGaP (accession number phs000128.v1.p1). The RA data was supported by NIH grant AR44422 and NIH contract N01-AR-7-2232. This work was partially supported by USA National Institute of Health grants AA01572, DA023166, GM070789, GM69590, HL087700, HL088215, HL088655, HL87660; USA National Science Foundation grant DMS 0714669; American Cancer Society grant IRG-58-010-50; a grant from the Urological Research Foundation, the German Federal Ministry of Education and Research BMBF (01GS0837); Fonds de Ia recherche en santé du Québec (FRSQ); and Fogarty International Postdoctoral Fellowship TW0511-05. The authors thank the GAW16 Group 9 participants for their stimulating discussion and for sharing their insights into the epistasis problem.

References

Alpaydin, E. Introduction to machine learning. Cambridge: MIT Press; 2004.

- Amos CI, Chen WV, Seldin MF, Remmers E, Taylor KE, Criswell LA, Lee AT, Plenge RM, Kastner DL, Gregersen PK. Data for Genetic Analysis Workshop 16 Problem 1, association analysis of rheumatoid arthritis data. BMC Proc. 2009; 3(Suppl 7):S2. [PubMed: 20018009]
- An P, Feitosa M, Ketkar S, Adelman A, Lin S, Borecki IB, Province MA. Epistatic interactions of *CDKN2B-TCF7L2* for risk of type 2 diabetes and *CDKN2B-JAZF1* for triglyceride/high-density lipoprotein ratio longitudinal change: Evidence from the Framingham Heart Study. BMC Proc. 2009; 3(Suppl 7):S71. [PubMed: 20018066]

Bateson W. Facts limiting the theory of heredity. Science. 1907; 26:649-60. [PubMed: 17796786]

- Borecki IB, Province MA. Genetic and genomic discovery using family studies. Circulation. 2008; 118:1057–63. [PubMed: 18765388]
- Breiman L. Random forests. Mach Learn. 2001; 45:5-32.
- Brunner, E.; Domhof, S.; Langer, F. Nonparametric analysis of longitudinal data in factorial experiments. New York: Wiley; 2001.
- Carlborg O, Hocking P, Burt D, Haley C. Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. Genet Res. 2004; 83:197–209. [PubMed: 15462413]
- Chanda P, Zhang A, Sucheston L, Ramanathan M. A two-stage search strategy for detecting multiple loci associated with rheumatoid arthritis. BMC Proc. 2009; 3(Suppl 7):S72. [PubMed: 20018067]
- Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. Am J Hum Genet. 2006; 79:1002–16. [PubMed: 17186459]
- Clarke GM, Pettersson FH, Morris AP. A comparison of case-only designs for detecting gene×gene interaction in rheumatoid arthritis using genome-wide case-control data in Genetic Analysis Workshop 16. BMC Proc. 2009; 3(Suppl 7):S73. [PubMed: 20018068]
- Cordell HJ. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet. 2002; 11:2463–8. [PubMed: 12351582]
- Culverhouse R, Klein T, Shannon W. Detecting epistatic interactions contributing to quantiative traits. Genetic Epidemiol. 2004; 27:141–52.
- Culverhouse R, Jin W, Jin CH, Hinrichs T, Suarez BK. Power and false-positive rates for the restricted partition methods (RPM) in a large candidate gene data set. BMC Proc. 2009; 3(Suppl 7):S74. [PubMed: 20018069]
- Cupples LA, Heard-Costa N, Lee M, Atwood LD. for the Framingham Heart Study Investigators. Genetics Analysis Workshop 16 Problem 2: The Framingham Heart Study data. BMC Proc. 2009; 3(Suppl 7):S3. [PubMed: 20018020]
- Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum Hered. 2008; 66:87–98. [PubMed: 18382088]

- Frankel WN, Schork NJ. Who's afraid of epistasis? Nat Genet. 1996; 14:371–3. [PubMed: 8944011]
- Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edin. 1918; 52:399–433.
- Greenland, S.; Rothman, KJ. Concepts of interaction. In: Rothman, KJ.; Greenland, S., editors. Modern epidemiology. 2. Philadelphia: Lippincott-Raven Publishers; 1998. p. 329-42.
- Greenland, S.; Lash, TL.; Rothman, KJ. Concepts of interaction. In: Rothman, KJ.; Greenland, S.; Lash, TL., editors. Modern epidemiology. 3. Philadelphia: Lippincott-Williams & Wilkins; 2008. p. 71-86.
- Huang C-H, Cong L, Xie J, Qiao B, Lo S-H, Zheng T. Rheumatoid arthritis-associated gene-gene interaction network for rheumatoid arthritis candidate genes. BMC Proc. 2009; 3(Suppl 7):S75. [PubMed: 20018070]
- Jung J, Song JJ, Kwon D. Allelic based gene-gene interactions in rheumatoid arthritis. BMC Proc. 2009; 3(Suppl 7):S76. [PubMed: 20018071]
- Kovac IP, Dubé M-P. Different models and single-nucleotide polymorphisms signal the simulated weak gene-gene interaction for a quantitative trait using haplotype-based and mixed models testing. BMC Proc. 2009; 3(Suppl 7):S77. [PubMed: 20018072]
- Kraja AT, Culverhouse R, Daw EW, Wu J, Van Brunt A, Province MA, Borecki IB. The Genetic Analysis Workshop 16 Problem 3: Simulation of heritable longitudinal cardiovascular phenotypes based on actual genome-wide single-nucleotide polymorphisms in the Framingham Heart Study. BMC Proc. 2009; 3(Suppl 7):S4. [PubMed: 20018031]
- Leamy LJ, Workman MS, Routman EJ, Cheverud JM. An epistatic genetic basis for fluctuating asymmetry of tooth size and shape in mice. Heredity. 2005; 94:316–25. [PubMed: 15674385]
- Li J, Tang R, Biernacka JM, de Andrade M. Identification of gene-gene interaction using principal components. BMC Proc. 2009; 3(Suppl 7):S78. [PubMed: 20018073]
- Liang X, Gao Y, Lam TK, Li Q, Falk C, Yang XR, Goldstein AM, Goldin LR. Identifying rheumatoid arthritis susceptibility genes using high-dimensional methods. BMC Proc. 2009; 3(Suppl 7):S79. [PubMed: 20018074]
- Lo SH, Zheng T. Backward haplotype transmission association (BHTA) algorithm a fast multiplemarker screening method. Hum Hered. 2002; 53:197–215. [PubMed: 12435884]
- Loh W-Y. Regression trees with unbiased variable selection and interaction detection. Stat Sinica. 2002; 12:361–86.
- Malzahn D, Balavarca Y, Lozano JP, Bickeböller H. Tests for candidate-gene interaction for longitudinal quantitative traits measured in a large cohort. BMC Proc. 2009; 3(Suppl 7):S80. [PubMed: 20018076]
- Manning AK, Ngwa JS, Hendricks AE, Liu C-T, Johnson AD, Dupuis J, Cupples LA. Incorporating biological knowledge in the search for gene×gene interaction in genome-wide association studies. BMC Proc. 2009; 3(Suppl 7):S81. [PubMed: 20018077]
- Mukherjee O, Sanapala KR, Anbazhagana P, Ghosh S. Evaluating epistatic interaction signals in complex traits using quantitative traits. BMC Proc. 2009; 3(Suppl 7):S82. [PubMed: 20018078]
- Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. Am J Epidemiol. 2009; 169:219–26. [PubMed: 19022827]
- Nelder J, Wedderburn R. Generalized linear models. J Roy Stat Soc Series A. 1972; 135:370-84.
- Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactordimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet. 2001; 69:138–47. [PubMed: 11404819]
- Therneau, T.; Grambsch, P. Extending the Cox model. New York: Springer-Verlag; 2000. Modeling survival data.
- Yao L, Zhong W, Zhang Z, Maenner MJ, Engelman CD. Classification tree for detection of singlenucleotide polymorphism (SNP)-by-SNP interactions related to heart disease: Framingham Heart Study. BMC Proc. 2009; 3(Suppl 7):S83. [PubMed: 20018079]
- Zheng T, Wang H, Lo SH. Backward genotype-trait association (BGTA)-based dissection of complex traits in case-control designs. Hum Hered. 2006; 62:196–212. [PubMed: 17114886]



Figure 1.

Gene×gene interactions found in RA GAW16 data set Problem 1 by various methods.



Figure 2. Dependence of the existence of epistasis on scale of response and model choice (link function)

In Figures 2a and 2b, we show the interaction between genes G and F in which the probability of disease (penetrance) for the baseline genotype group is P disease | G=aa, F=bb]=0.001, and each A allele dose for gene G increases the risk by six-fold while each B allele for gene F increases the risk by five-fold. In Figure 2a, both genes show no dominance and no interaction on the multiplicative probability scale (all three genotype lines are linear and parallel), which would be the conclusion from a log-linear model, whereas these same data show strong dominance (non-linear response by genotype) as well as strong G×G interaction (non-parallel lines) on the log-odds scale, which would be the conclusion from logistic regression. By contrast, in Figures 2c and 2d, we show two other genes J and K in which the baseline genotype group penetrance is P[disease | J=cc, K=dd] = 0.5, (which corresponds to odds(P)=1), and each Callele dose for gene J increases the odds four-fold, while each D allele dose for gene K increases the odds two-fold. Here, both genes show strong dominance as well as G×G interaction on the multiplicative probability scale (Figure 2c) which would be the conclusion from a log-linear model, but these same data show no dominance and no interaction on the log(odds) scale for logistic regression (Figure 2d). Hence, the existence or lack of epistasis and/or dominance is dependent upon the scale of the response and therefore also on the choice of model (link function).

An et al.



Figure 3. Examples of epistasis which persist on every scale

Epistasis for mouse molar size (a), mouse molar shape (b) [reproduced from Leamy et al., 2005], and for chicken body weight at 6 weeks of age (c) [reproduced from Carlborg et al., 2004].

Table I

Methods to detect and estimate G×G Interactions utilized in GAW16

Group 9 contribution	Method	Idea	Method Ref
	Statistical	Most based upon GLM $\mathbf{E}[\mathbf{Y} \mathbf{X}] = \mathbf{I}^{-1}(\mathbf{D}[\mathbf{X}] \mathbf{\beta})$ where <i>X</i> are predictors; $\{G_{j}, G_{j}\} \subseteq X$ for gene variants $G_{j}, G_{j}; [G_{i} G_{j} G_{j} \times G_{j}] \subseteq D[\mathbf{X}]$ (design matrix contains both main effects as well as interactions between G_{i} and G_{j}) and $I(.)$ is link function depending upon the scale of <i>Y</i> (identity, logistic; proportional odds; multinomial, proportional hazards, etc.)	
Clarke et al.	Case-only	$\mathbf{E}[G_i G_j] = 1^{-1} \left(\mathbf{D}[G_j] \mathbf{\beta} \right) \text{ (test if } G_j \text{ predicts } G_i \text{ in cases)}$	
An et al.	Family mixed model	E [Y X , PEDID] = l^{-1} (D [X] β) + D [PEDID] γ) where $\gamma \sim N(0, \Sigma)$ (add random effect, γ , for pedigrees)	
Kovac et al. Jung et al.	Allelic score	E [Y X] = 1 ⁻¹ (D [X] β) $S(A_{iμ}, A_{jν}) ⊆ D[G × G]$ for $u, v = 1, 2$, where $S(A_{iμ}, A_{jν})$ $= P[\langle A_{iμ} \rangle \langle A_{jν} \rangle G_i = A_{i1}A_{i2}, G_j = A_{j1}A_{j2}] = P[\langle A_{iμ} \rangle G_i = A_{i1}A_{j2}] P[\langle A_{jν} \rangle G_i = A_{j1}A_{j2}]$, where $\langle A \rangle$ is the state value of allele A (assuming independent loci <i>i</i> , <i>j</i>)	
Liang et al.	Omnibus test	Form L(G) latent factors from linear combinations of G loci, and test E [Y L (G)] = 1^{-1} (D [L (G)] β) with $L_m \times L_n$ latent interactions; infer $G_i \times G_j$ interactions from $L_m \times L_n$ interactions and latent path loadings	Chatterjee et al., 2006
Li et al.	PCA	Form PCA(G) PCs of <i>G</i> loci, and test E [Y PCA (G)] = \mathbf{l}^{-1} (D [PCA (G)] β) with <i>PCA_m×PCA_n</i> interactions; infer <i>G_i×G_j</i> interactions from <i>PCA_m×PCA_n</i> interactions and factor loadings	Murcray et al., 2009
Manning et al.	Interactions among significant main effect genes	Test $\beta[G_i \times G_j] = 0$ terms only for significant $\beta[G_i] = 0$ or $\beta[G_j] = 0$ main effects	
Manning et al.	Interactions among pathway genes	Test $\beta[G_i \times G_j] = 0$ terms only when $\{G_i, G_j\} \subseteq$ Pathway	
Li et al.	Two-step approach	Test $\beta[G_i \times G_j]$ = 0 terms only when GLM $E[G_i \mid G_j]$ = l^{-1} (G_j $\beta)$ significant	Murcray et al., 2009
Malzahn et al.	LNPT	No GLM, longitudinal rank-sum test , no assumptions about distribution F of quantitative trait, test contrast between trait distributions F^{kl} of two-locus genotype strata (k, l)	Brunner et al., 2001
Malzahn et al.	Proportional hazards model	Event time as age at first exam when quantitative longitudinal trait crossed a predefined threshold, multiplicative hazard ratio null H_0 : $HR(G_1, G_2)=HR(G_1)HR(G_2)$.	Therneau and Grambsch, 2000
Kovac et al.	UNPHASED	Combine genotypes into estimated haplotypes; <i>H</i> , to test $E[Y X] = 1^{-1} (D[X] \beta)$, where $H \subseteq X$	Dudbridge, 2008
	Machine Learning	High dimensional heuristic search algorithms to detect G×G interactions that mostly rely upon split samples with cross validation to avoid fitting to noise. Some use the basic GLM to evaluate particular interactions, but many of these detection methods do not utilize a formal G×G model per se, and therefore do not provide an estimate of effect size. The emphasis is on efficient search among a large number of possible G×G interactions to determine which are signals and which are noise, rather than on detailed modeling of any particular G×G interaction.	
Liang et al.	MDR	$ \forall \text{ pair of genes } G_{j_i} \ G_{j_j} \text{ form binary } \mathbf{I}\{\mathbf{g}_i \times \mathbf{g}_j \mid \mathbf{g}_i \in \mathbf{G}_i, \mathbf{g}_j \in \mathbf{G}_j, \mathbf{O}(\mathbf{Y}, \mathbf{g}_i \times \mathbf{g}_j) > 1\} \text{ and fit GLM, where O}(Y, G_{j'} > G_{j}) = E[Y=1 \mid G_i \times G_j] / E[Y=0 \mid G_i \times G_j] $	Ritchie et al., 2001
Mukherjee et al.	GMDR	\forall pair of genes G_h , G_f form binary $\mathbf{I}\{\mathbf{g}_i \times \mathbf{g}_j \mathbf{g}_i \in \mathbf{G}_i, \mathbf{g}_j \in \mathbf{G}_j, \mathbf{S}(\mathbf{Y}, \mathbf{g}_i \times \mathbf{g}_j) > T\}$ and fit GLM, where $S(Y, g_i \times g_j)$ is a score function measuring relationship between <i>Y</i> and $G_i \times G_j$ and <i>T</i> is threshold	

An et al.

Group 9 contribution	Method	Idea	Method Ref
Culverhouse et al.	RPM	Pool { G_i, G_j } whenever $ \mathbf{E}[\mathbf{Y} \mid \mathbf{G_i}] \mathbf{E}[\mathbf{Y} \mid \mathbf{G_j}] < \epsilon$ and fit GLM	Culverhouse et al., 2004
Liang et al.	RF	Bootstrap samples of subjects and G_i values to fit TREEs (recursively partition sample by binary splits $I[G_i \ c]$ to maximize Y purity)	Breiman, 2001
Yao et al.	GUIDE	TREE model which selects variables to split nodes via chi-square test	Loh, 2002
Chanda et al.	KWII	Stepwise select gene <i>G</i> into <i>S</i> that gives largest KWII (<i>S</i>) = $\Sigma_{t \subseteq S}$ (1) ^{S - t} <i>H</i> (<i>t</i>) that is not redundant to <i>S</i> or <i>G</i> [where <i>H</i> (<i>t</i>) = $\Sigma_{t \in u}$ <i>P</i> (<i>t</i> = <i>u</i>) log ₂ (<i>P</i> (<i>t</i> = <i>u</i>)) is Shannon entropy information]	
Huang et al.	GTD	$V = \sum_{g \in G} (E[Y=1 g] - E[Y=0 g])^2$ to measure joint effects of SNPs interaction of SNPs are measured by excess effect of a pair over their marginal effects; assess significance by permutation test	Lo and Zheng, 2002; Zheng et al., 2006

Table II

Common GLM Examples

P (Probability distribution)	<i>l</i> (Link function)	Corresponds to this type of model
Normal	Identity: µ	Linear regression
Binomial	Logit: logit: log $(\pi/(1-\pi))$	Logistic regression
Binomial	Probit: $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mu} \exp(-\frac{1}{2}Z^2) dZ$	Probit regression
Poisson	Log: log(µ)	Poisson regression
Poisson	Log-hazard ratio: $log(h(t)/h_0(t))$	Proportional hazards
Negative-binomial	Log: $log(\mu)$	Negative-binomial regression
Gamma	Reciprocal: 1/µ	Gamma regression