Title: Contrasting LD patterns between cases and controls as a novel association mapping method.

(accepted to *American Journal of Human Genetics*)

Authors: Dmitri V. Zaykin[♯], Zhaoling Meng[†], Margaret G. Ehm[‡],

[♯] National Institute of Environmental Health Sciences, National Institutes of Health.

[†] Biostatistics & Programming, Sanofi-Aventis.

[‡] Genetics Research, GlaxoSmithKline Inc.

Running head: Standardized LD for association mapping

Corresponding author:

Dmitri Zaykin (email: zaykind@niehs.nih.gov)

National Institute of Environmental Health Sciences

MD A3-03, South Bldg (101)/B356B POB 12233

Research Triangle Park, NC 27709

Phone: (919) 541–0096; Fax: (919) 541–4311

Abstract

Identification and description of genetic variation underlying disease susceptibility, efficacy and adverse reactions to drugs remains a difficult problem. One of the important steps in the analysis of variation in a candidate region is the characterization of linkage disequilibrium (LD). In a region of genetic association, the extent of LD varies between the case and the control groups. Separate plots of pairwise standardized measures of LD (e.g. D'), for cases and controls are often presented for a candidate region to graphically convey case-control differences in LD. However, the observed graphical differences lack statistical support. Therefore we suggest the "LD contrast" test to compare whole matrices of disequilibrium between two samples. A common technique of assessing LD when the haplotype phase is unobserved is the EM algorithm with the likelihood incorporating the assumption of Hardy-Weinberg equilibrium (HWE). This approach presents a potential problem in that in the region of genetic association the HWE assumption may not hold when samples are selected on the basis of phenotypes. Here we present a computationally feasible approach that does not assume HWE, along with graphical displays and a statistical comparison of pairwise matrices of LD between case and control samples. LD contrast tests provide a useful addition to existing tools of finding and characterizing genetic associations. Although haplotype association tests are expected to provide superior power when susceptibilities are primarily determined by haplotypes, the LD contrast tests demonstrate substantially higher power under certain haplotype-driven disease models.

Introduction

There has been considerable progress in designing techniques that go beyond sequential testing of single nucleotide polymorphisms (SNPs). These methods are particularly important for the analysis of multiple SNPs that jointly represent variation within common transcripts and other functional regions such as promoters. Methods for detection of association between traits and interacting genetic polymorphisms are being rapidly developed. Many approaches have considered important situations where haplotypes of consecutive markers can be defined and tested for association with the trait. Methods have been designed to incorporate various sampling designs as well as the haplotype phase uncertainty[1,2,3,4,5,6,7]. It has been noted that the extent of LD can be different between the case and the control groups in a region of genetic association, and the case-control LD comparison can aid the analysis in a region of putative association[8]. Contrasting pairwise LD matrices between cases and controls via graphical display provides a direct visual comparison[9]. However, the observed graphical difference is subject to sampling variation and lacks statistical support. Therefore, a statistical test is desirable. In the context of association mapping, Nielsen *et al.* presented a direct LD comparison approach involving two diallelic loci and noted that in certain situations a test that directly compares LD extent between the case and the control groups can be a powerful alternative to either haplotype-based or single-marker approaches[10]. A test comparing the LD extent will include only a single LD parameter that results in a single degree of freedom test, whereas

a haplotype test will include four haplotypes with three degrees of freedom. Nielsen *et al.* considered the case of unambiguous haplotype phase. When the haplotype phase is unknown, the EM algorithm is used to infer frequencies of haplotypes and ultimately to assess LD. The likelihood is constructed by assuming HWE on the level of haplotypes. With two diallelic markers there are four haplotypes, and the usual assumption is that the two-locus haplotypes are in HWE. Checking that each SNP is in HWE is not sufficient to ensure HWE at the haplotypic level. Furthermore, in the region of association, the HWE is generally expected to be distorted in case and control samples[11,12]. Therefore, the EM computation, while being a valuable tool for evaluating LD in a sample of population controls, is not strictly appropriate for comparing LD levels in samples of cases and controls, or samples otherwise selected on the basis of phenotype.

Recently, Schaid[13] and Zaykin[14] showed that LD estimation using the composite disequilibrium approach, discussed below, provides similar results to the EM-based method under HWE, is computationally simpler, and avoids the assumption of haplotypic HWE. Hamilton and Cole[15] and Zaykin[14] gave bounds and proposed normalization for LD based on the composite definition. Zaykin showed that this normalization is robust to departures from HWE. Therefore, we propose using the composite coefficient and its normalization for characterizing the LD in case-control samples. This leads to efficient methods of comparing and testing the difference of pairwise LD matrices between the case and control samples. We show that certain disease models result in high power of the LD contrast test in comparison with the haplotypic test even under situations when susceptibilities are largely

determined by haplotypes. In such situations the LD contrast test outperforms both the haplotype-based test as well as multilocus tests based on comparison of SNP scores.

Methods

A test for comparing two single linkage disequilibrium (LD) coefficients ("LD contrast"

test) was described by Nielsen *et al.* for the case of known phase, *i.e.* when four haplotype

classes are directly observed[10]. A chi-square test statistic has the form

$$X^2 = \frac{\left(\hat{D}_{AB}^{\text{cases}} - \hat{D}_{AB}^{\text{controls}}\right)^2}{\text{Var}\left(\hat{D}_{AB}^{\text{cases}}\right) + \text{Var}\left(\hat{D}_{AB}^{\text{controls}}\right)} \tag{1}$$

with the variances given in Weir[16], p. 113. The LD coefficient, D_{AB}, is equal to

$P_{AB} - p_A p_B$, P_{AB} is the frequency of haplotype carrying alleles A, B, and p_A, p_B are the

corresponding allele frequencies. A log-linear framework for comparing disequilibria

coefficients at one and two loci has been provided by Huttley and Wilson[17]. There is only a

single LD parameter describing dependence among the four haplotypes and the

corresponding allele frequencies. Therefore, this LD contrast test has an advantage in

being a single degree of freedom test, whereas a haplotypic test with all four haplotypes

has three degrees of freedom. Nielsen *et al.* found, empirically, that this test can have

higher power than either a haplotypic or a single-marker test when the pairwise LD

between a functional site and two tested markers is low. When the haplotype phase is

unknown, the above test can be extended to compare two composite LD coefficients using a

test analogous to (1) with the variances provided by Weir[16]. Furthermore, comparison of

standardized coefficients may be of interest when single locus genotypic frequencies differ.

One of the commonly used standardized measures of LD is the coefficient D'_{AB} suggested

by Lewontin[18], $D'_{AB} = D_{AB}/\max D_{AB}$, where $\max D_{AB}$ is the maximum possible absolute

value of D_{AB} given allele frequencies, also described in a non-genetic context by Yule[19],

$$\max D_{AB} = \begin{cases} \min\left[p_A p_B, \ (1-p_A)(1-p_B)\right]; & \text{if } D_{AB} < 0 \\ \\ \min\left[(1-p_A)p_B, \ p_A(1-p_B)\right]; & \text{if } D_{AB} > 0. \end{cases} \tag{2}$$

Weir[20] discussed the correlation between alleles A, B given by

$\mathcal{R}_{AB} = D_{AB}/\sqrt{p_A(1-p_A)p_B(1-p_B)}$, which has the range that depends on allele

frequencies. Weir[20] and Peduzzi $et\ al.$[21] gave the bounds for r_{AB}. While allele counts are

directly observed, haplotype phase is often ambiguous, therefore P_{AB} cannot be estimated

as a proportion of AB haplotypes among all $2n$ haplotypes in a sample. The maximum

likelihood estimate $\hat{P}_{AB}$, and correspondingly, $\hat{D}_{AB}$ and $\hat{D}'_{AB}$ can be obtained. However,

this approach usually requires the assumption of HWE, that is the dilocus genotype

frequencies are given by the products of frequencies of haplotypes. Weir and Cockerham[22]

suggested estimating the composite LD coefficient instead, defined as

$\Delta_{AB} = P_{AB} + P_{A/B} - 2p_A p_B$, with the composite correlation

$$r_{AB} = \frac{\Delta_{AB}}{\sqrt{(p_A(1-p_A) + D_A)(p_B(1-p_B) + D_B)}} \tag{3}$$

where D_A, D_B are the the Hardy-Weinberg disequilibrium coefficients at two loci, and $P_{A/B}$

is the joint frequency of alleles A and B at two different gametes. This coefficient is directly

estimated from dilocus counts[16], and under HWE corresponds to D_{AB}. Weir[20] and Schaid[13]

investigated statistical properties of the composite LD estimator and made comparisons of

the composite ($\hat{\Delta}_{AB}$) and the maximum likelihood ($\hat{D}_{AB}$) estimators. The composite

estimator appears to perform well, being robust with respect to the HWE assumption.

The maximum and minimum possible values for Δ_{AB} given genotypic frequencies at two loci were reported by Hamilton and Cole[15] and Zaykin[14]. These values correspond to bounds on covariance between two trinary variables that take values $-1, 0, 1$. Equation (4) in Zaykin[14] gives the bounds for abs(Δ_{AB}). succinctly as

$$\left.\begin{aligned} &\max \Delta_{AB} = \tfrac{d-s}{2} + \tfrac{1}{2}(1 - 2p_A)(1 - 2p_B) \\ &\text{where } d = \min(P_{AA}, P_{bb}) + \min(P_{aa}, P_{BB}) \end{aligned}\right\}, \quad \Delta_{AB} < 0$$

$$\left.\begin{aligned} &\max \Delta_{AB} = \tfrac{d-s}{2} - \tfrac{1}{2}(1 - 2p_A)(1 - 2p_B) \\ &\text{where } d = \min(P_{AA}, P_{BB}) + \min(P_{aa}, P_{bb}) \end{aligned}\right\}, \quad \Delta_{AB} > 0$$

$$\text{and } s = 1 - d - \min(1 - d, P_{Aa} + P_{Bb}) \tag{4}$$

The standardized composite measure of LD with the range -1 to 1 is computed as

$$\Delta'_{AB} = \frac{\Delta_{AB}}{\max \Delta_{AB}}. \tag{5}$$

The standardization using (4) takes into account composite LD dependency on genotype frequencies and holds the promise for association mapping applications. Cases and controls will generally have different extent of gametic as well as non-gametic disequilibria around a region of genetic association[10,14], which is captured by the composite LD.

Such a test is extended here to the comparison of whole matrices of standardized coefficients between the case and the control groups to aid in identification of effects due to interactions among SNPs. The matrix of non-standardized $\hat{\Delta}$ coefficients is non-negative definite by virtue of being a variance-covariance matrix. Therefore one can compute

statistics based on eigenvalue-eigenvector (spectral) decomposition of the LD matrix.
Previously, we proposed the use of spectral decomposition of the composite LD matrix for
selection of a subset of markers that optimize the information retained in a genomic region,
using samples of population controls[23]. The matrix of standardized composite LD is not
necessarily positive definite, limiting the application of the spectral decomposition-based
statistic. Another statistic used in this study is based on the overall LD difference (Z_2).

We define the standardized composite LD matrices as $\mathbf{\Delta}'_Y$ and $\mathbf{\Delta}'_N$ and matrices of the
composite LD correlation (3) for the case and the control groups as $\mathbf{r}_Y$ and $\mathbf{r}_N$. In both
cases, diagonal entries of the matrices are equal to one. Composite LD matrices have
spectral decompositions

$$\hat{\mathbf{r}}_Y = \sum_{i=1}^{L} \lambda_{i|Y} \, \mathbf{e}_{i|Y} \, \mathbf{e}_{i|Y}^T \tag{6}$$

$$\hat{\mathbf{r}}_N = \sum_{i=1}^{L} \lambda_{i|N} \, \mathbf{e}_{i|N} \, \mathbf{e}_{i|N}^T \tag{7}$$

where $\left\{ \lambda_{i|N}, \mathbf{e}_{i|N} \right\}$ are sample composite LD eigenvalues and eigenvectors (for the control,
N, or for the case, Y, group), and T denotes the transpose. Spectral decompositions based
on the composite LD covariance matrices, $\mathbf{\Delta}_Y, \mathbf{\Delta}_N$, are defined similarly.

We define matrices of first k column case and control eigenvectors by $\mathbf{E}_Y$ and $\mathbf{E}_N$,
correspondingly. The two statistics are

$$Z_1 = \text{trace} \left[\mathbf{E}_Y^T \mathbf{E}_N \mathbf{E}_N^T \mathbf{E}_Y \right] \tag{8}$$

$$Z_2 = \text{trace} \left[(\hat{\mathbf{r}}_Y - \hat{\mathbf{r}}_N)^T (\hat{\mathbf{r}}_Y - \hat{\mathbf{r}}_N) \right] \tag{9}$$

We suggest that the Z_2 statistic should take a slightly different form when computed using

the standardized LD:

$$Z_2 \;=\; \frac{\mathrm{trace}\left[(\boldsymbol{\Delta}_Y' - \boldsymbol{\Delta}_N')^T(\boldsymbol{\Delta}_Y' - \boldsymbol{\Delta}_N')\right]}{4L(L-1)} \tag{10}$$

In these equations, L is the number of markers, and $k \leq L$ is the number of principal

components. The denominator, $4L(L-1)$, is the upper bound for the numerator of Z_2.

The denominator does not affect the magnitude of the resulting p-value because it is

invariant under permutations.

The statistic Z_1 measures the difference between two spaces (sum of squared cosines of

the angles between the eigenvectors) defined by the first k eigenvectors and ranges between

k and 0 (maximum difference). Krzanowski described this statistic and the corresponding

permutation-based tests (where the phenotype value is randomly shuffled among

individuals) for comparison of two sets of principal components[24,25]. The value k must be

specified in advance. Krzanowski[25] suggested using the value of k that is the largest integer

smaller than $L/2$. This ensures that the "important" components are represented, while

values $k \geq L/2$ will cause the subspaces defined by the two sets of eigenvectors to intersect

in at least one dimension.

The sum of squared differences statistic Z_2 measures the overall difference in the

corresponding pairwise LD. This statistic is also appropriate for comparing Δ_Y' and Δ_N'.

Note that for the standardized LD, the range of Z_2 is ($0 \leq Z_2 \leq 1$) with 1 giving the

maximum difference.

Both Z_1 and Z_2 definitions can be covariance based, rather than correlation based.

However, results for the covariance based tests are not reported here because these tests showed consistently inferior power when compared to the correlation based tests. We also performed a preliminary examination of several statistics based specifically on the comparison of corresponding LD eigenvectors, as well as eigenvalues (e.g. sum of squared differences) between the case and the control groups. These tests did not show prominent power characteristics and the results are not reported here. Nevertheless, detailed study of utility of such tests may warrant further investigation.

The generalized T^2 test was applied in the association mapping context by Xiong *et al.*[26] This test employs the composite LD matrix as part of the test statistic. The T^2 test compares mean vectors of SNP values in cases and controls where SNP values are obtained by recoding genotypes as $AA \rightarrow 1$, $Aa \rightarrow 0$, $aa \rightarrow -1$. The variance part of the T^2 test statistic is the pooled variance-covariance matrix for the recoded values. It follows that under the hypothesis of no association the off-diagonal elements of this matrix are estimates of twice the composite LD coefficients, and the diagonal entries are twice the estimates of the variances of allele frequencies. Therefore, the generalized T^2 test indirectly utilizes the composite LD in the variance part of the statistic.

Results

To evaluate performance of the proposed tests, we compared methods that are designed to detect either single-SNP effects, or SNP interactions when the effects are associated with entire haplotypes. The T^2 test is expected to have good power in the presence of several SNPs contributing to the association. In contrast, the $\min P$ test[27] is most sensitive to a single associated SNP while accounting for correlation between SNPs due to LD. This test evaluates significance of the most extreme association test statistic (Armitage's trend test in the present study). The significance is evaluated via permutations, preserving dependencies among SNPs. To detect haplotypic effects we employed the "Haplotype Trend Regression" method (HTR) of Zaykin *et al.*[2] Methods used for power comparisons in this study are merely providing a reference point of comparison under different models. It is unlikely that a single "best" method can be recommended for the discovery of genetic associations because the power obtained for the different methods will vary with the disease models assumed.

Pharmacogenetic association mapping example – CYP2D6

Identification of individual genetic differences in response to medicine has potential for reducing side effects and improving efficacy of drugs. The cytochrome p450 gene, CYP2D6, is involved in metabolism of about 20% of marketed drugs[28]. Hosking *et al.*[29] described the association of single nucleotide and haplotype polymorphisms with the poor drug

metabolizer phenotype in a region around the CYP2D6 gene. The data set consisted of 41 "poor metabolizer" cases and 977 controls. SNPs from the middle of the region show very high levels of association that would be strongly supported by any of the tests discussed here. To illustrate an application of our technique, we identified six 5'-flanking consecutive SNPs. Missing genotypes were imputed with the package MICE[30]. Further details of the data set are given in Hosking *et al.*

We found pronounced differences in LD between the case and the control groups. Figure 1 presents the differences in LD graphically, by displaying the LD matrices using ellipses whose shape reflects the magnitude of LD and the direction reflects the sign of the disequilibrium: 45°-oriented ellipses reflect the positive sign of LD, while round shape of an ellipse reflects a low degree of LD. Murdoch and Chow[31] suggested the usage of such graphs to display correlation matrices. Evidently, there are large observed differences, as some of the coefficients are reversed in sign. The values of r (the left graph) and Δ' (the right graph) are similar to each other.

The difference in correlation is significant at the 5% level: the p-value for the $Z_2(r)$ test is equal to 0.033, although the $Z_2(\Delta')$ test p-value of 0.061 does not reach significance (all tests except the asymptotic T^2 are based on 50,000 permutations). The test comparing the first two correlation-based principal components, $Z_1(k=2)$, gave a significant p-value, 0.026. Statistics based on $k = 1, 3$ resulted in p-values equal to 0.232 and 0.283, correspondingly. There is a multiple testing issue involved with trying out statistics based on the different numbers of principal components, $k = 1, 2, 3$. Nevertheless, we note that

the value $k = 2$ corresponds to Krzanowski's recommendation, and could be set as the default value. The T^2 test gave the p-value equal to 0.337, reflecting the apparent absence of detectable effects associated with individual SNPs. Neither the allelic trend test, nor the test comparing genotypic frequencies at individual SNP were significant. The overall haplotypic test was not significant (p=0.168). Thus, the application of the LD contrast test to this particular data set shows that the method is successful in detecting the case-control LD difference, supporting visual differences in the LD patterns conveyed by Figure 1.

Simulation I: 5-SNP and 6-SNP haplotypes

A more extensive evaluation of the tests based on the Z_1 and Z_2 statistics was performed using simulations. When susceptibilities are mainly driven by haplotypes (i.e. there are pronounced haplotype effects but no interaction between haplotypes), it is expected that haplotypic tests should have optimal power. Nevertheless, there are notable exceptions from this rule. For two markers, Nielsen et $al.$[10] showed that there are scenarios when a test comparing LD coefficients is more powerful than a single-locus or a haplotypic test. One situation when this is the case is when multilocus susceptibilities induce an "orthogonal"-like distribution of dilocus haplotypes between the case and the control groups. By "orthogonal" we mean the situation when high-susceptibility haplotypes tend to be defined by different SNPs. Culverhouse et $al.$[32] considered epistatic models of this type.

   To mimic this scenario, a set of simulations has been constructed under a haplotype-driven model, common for all simulations. Haplotype frequencies were drawn

from the Dirichlet(1,...,1) distribution. Effect sizes were drawn from the Gamma(1)

distribution and inspected to ensure that two large effect sizes (h_i) are allocated to the

most distinct 6-SNP haplotypes, 111111 and 222222, corresponding to a situation of two

independent mutations in high LD with two very distinct haplotypes. To form an

individual, a pair of haplotypes in these simulations was sampled from the population with

the Dirichlet-derived haplotype frequencies. To obtain the binary outcome, the continuous

phenotype values $(Y_{ijk} = h_i + h_j + e_k)$, where $e_k \sim N(0, \sigma^2)$, $\sigma = 7.5$, were dichotomized

around two different threshold values, determined by the 0.05 and 0.5 population quantiles

of Y. The population values of Δ' among the cases and among the controls are listed in

Table 1. The correlation LD values followed the same pattern and were similar in values to

the values of Δ'. The largest difference between the corresponding r and Δ' coefficients was

found to be 0.06. The population LD values were small, which may correspond to a

situation when a set of SNPs in a candidate gene is selected on the basis of redundancy

reduction[23]. The largest case-control LD difference (0.116) was between the (1,4) and (4,1)

entries of the LD matrix, the minimum difference was 0.009, and the mean difference was

0.07. For this set of simulations, 250 cases and 250 controls were sampled for each of 10000

simulation runs.

Tests based on Z_1 and Z_2 statistics (using both Δ' and r-based versions of Z_2) were

performed and p-values recorded. Results of these simulations for the two values of

population prevalence are shown in Table 2. The results show that the LD-contrast test

based on the squared difference statistic (Z_2) has the largest power using both the Δ' and

the correlation based definitions.

The power of the haplotype-based test[2] was substantially lower, and the power of both T^2 and $\min P$ (single-SNP permutation-based trend allelic test) was low. The power of the principal components based test (Z_1) was lower then the power of the test based on the Z_2 statistic. However, in this model it was higher than the power of the T^2, $\min P$, and the haplotypic tests (at the value of $k < L/2 = 2$). Dichotomization around the population mean to produce the binary outcome yielded similar results to the quantile-defined thresholds just described (data not shown). In addition to these results with fixed parameters, we conducted a set of 5-SNP simulations where samples of haplotypes were obtained using the forward evolutionary model of drift with recombination[33]. The simulations are "forward" to distinguish them from a popular coalescent approximation to this process, that operates "backwards" in time. These forward simulations are a typical implementation of a genetic drift with admixture population-genetic model, with non-overlapping generations and recombination modeled as a Poisson process. A very similar model was used in Zaykin et $al.$[2] The effects that determine susceptibilities were sampled from a template that induces pairwise orthogonality, with added normal variability. In contrast to the simulation just described, all population parameters were sampled anew prior to each simulation. This allowed averaging across a variety of models. The power is not necessarily expected to be reduced in this setup. In general, larger variance associated with haplotype effects would result in higher power values of the tests. In addition, the induced "marginal effects" at the level of SNPs and dilocus haplotypes are

dependent on both, the susceptibility values as well as on the population frequencies.

In these simulations, we observed that the 5×5 composite LD matrix comparison tests (Z_2) still had higher power on average (88% power for the correlation and 75% for Δ') than either the generalized T^2 (55 % power) or the haplotype-specific test (60% power). Thus, these simulations confirmed that the power of the LD contrast tests is still the highest, as was found to be the case for the fixed set of effects and frequencies.

Simulation II: 15 and 30-SNP haplotypes

An evaluation of the tests when the trait variation is determined by the diploid pairs of haplotypes (diplotypes) was performed using simulations. For this model we used much larger, 15-SNP and 30-SNP haplotypes sampled from a population generated by the forward evolutionary model of drift with recombination[33]. The phenotype model was similar to the one described above. We considered a more general, diplotype-driven model, where normally distributed diplotype rather than haplotype effects were added to the trait value together with the common normal error. New diplotype effects were sampled prior to each simulation. In this set of simulations the trait values have been dichotomized around the mean to produce a binary trait. The LD contrast tests were verified to have the correct type-I error by setting the population genetic effects to zero and examining quantiles of the resulting p-value distribution.

This set of simulations generated relatively high pairwise LD. The two middle quartiles for the population LD distribution (measured by r_{AB}) were estimated to be 0.413 and

0.975 with the median value 0.735. One of the 30-SNP samples from this simulation study was used to produce an illustrative graphical plot of pairwise LD (Figure 2). The plot illustrates LD differences between the upper (cases) and the lower (controls) samples. For example, there is a region of high LD around the marker pair (21,8) in the cases, while this region has relatively low LD in the controls. Nonetheless, statistical tests, as described here, are needed to assess the extent to which these LD differences can be attributed to the sampling variation.

As before, we assumed haplotypic phase to be unknown. Many published haplotype association mapping algorithms would not be computationally feasible given the large number of SNPs. The generalized T^2 test[26] has been used for the comparison as well as the single SNP based "min$P$" shuffling test, where the significance of the allelic trend test with the maximum value of chi-square is obtained via permutations[27,34]. It should be noted that the T^2 test has high power when alleles of multiple SNPs independently contribute to the trait, because the test compares means of SNP scores between the case and the control groups. In both 15-SNP and 30-SNP settings we observed similar power for the T^2 and Z_2 tests.

For the 15-SNP data, the power was 0.71 for T^2 and Z_2, when Z_2 was based on the correlation LD matrix (Table 3). The power for the Z_2 test based on the standardized matrix was lower, 0.62. The single-SNP permutation-based trend (allelic) test had power 0.57. Thus, despite taking into account the correlation between SNPs, single-marker tests had relatively lower power. We computed the eigenvector statistic Z_1 for values of k

required to account for various proportions of the variance (0.5, 0.75, 0.9) as can be determined by the cumulative sum of eigenvalues (such an approach has been previously employed in Meng *et al.*[23] The maximum value of k was set to 7. This resulted in k equal to 1-3 ($\bar{k} = 1.27$) for 0.5 of the variance, k equal to 1-6 ($\bar{k} = 3.97$) for 0.75 of the variance, and k equal to 6-7 ($\bar{k} = 6.99$) for 0.9 of the variance. Fixed values of k have been tried as well, however we could not achieve power comparable to that of the test based on Z_2. Due to higher LD in this set of simulations, the best power was observed at intermediate values of k.

Similar relative power was observed for the 30 SNP data. When Z_2 was based on the correlation LD matrix, the power was 0.82 for T^2 test, 0.85 for the Z_2 test, and 0.77 for the Z_2 test based on the standardized LD ($\mathbf{\Delta}'$). The eigenvector statistic based test (Z_1) had highest power of 0.62 at the proportion of the variance equal to 0.75, that corresponded to $k = 2\text{-}8$, ($\bar{k} = 5.71$). The single-SNP permutation-based trend test had very low power, 0.17.

Although the LD matrix comparison test was found to have power similar to the generalized T^2, our results suggest that these tests tend to identify essentially different attributes of genetic association in a region. The left graph of Figure 3 shows a plot of p-values obtained from the T^2 test versus the corresponding p-values of Z_2 (15 SNPs simulation). The correlation between the two tests was quite low (0.36) and over a half of T^2 test p-values greater than 0.05 were smaller than 0.05 when evaluated with Z_2. On the other hand, the right graph for the correspondence between Z_1 and Z_2 statistics shows very

large correlation (0.91). The test Z_1 for the difference between the case and the control

group principal components had lower power than the test based on Z_2, shifting points up

from the diagonal on the second graph.

Discussion

Genetic association studies typically report characterization of LD in candidate regions with LD plots, i.e. using graphical representation of LD matrices[9]. These plots are usually given for population control samples, although LD plots for case samples are reported as well and compared (visually) with the LD pattern in control samples. Rubio *et al.*[35] compared graphical plots of LD between multiple sclerosis case and control samples in the HLA region concluding that D' values appeared slightly higher in the case sample. Suarez *et al.*[36] compared specifically composite LD coefficients between samples of alcoholics and non-alcoholics, concluding that there was similarity in the pattern of LD, "although there is the suggestion of less disequilibria in the alcoholic sample than among the controls". We suggest that such comparisons should be complemented by a statistical procedure. Moreover, we found that the power of comparing LD patterns is comparable to that of traditional mapping techniques, or even superior in certain situations.

The standardized LD coefficient, D' remains a popular measure that accounts for dependence of the LD range on allele frequencies. However, one of the problems with a EM-based estimator is the requirement of the random union of haplotypes (haplotypic HWE). We resolve this problem by accommodating results of Hamilton and Cole[15] and Zaykin[14], and suggest that the plots can be based on the standardized composite coefficient. Straightforward definitions of the composite and standardized composite LD as well as efficiency of the computations make it easy to compare LD plots between samples of

cases and controls. Such comparisons can be done "by eye", but a statistical procedure is desirable. An asymptotic test to compare two LD (i.e. covariance) matrices can be constructed, however such tests are rather sensitive to distributional assumptions[37,38]. In addition, the distribution theory and inference become much more complicated once normalizations of covariance (e.g. correlation) are considered[25]. Because of these concerns, we adopted the permutational framework to provide comparison of LD matrices based on the composite coefficient and its standardized version.

Statistical approaches specifically tailored to identification of haplotype effects are being rapidly developed[39]. There is strong biological evidence that the entire haplotypes rather than single SNPs are important in determining the trait variation. Therefore identification and estimation of haplotype effects is an important issue. Still, the multiplicity of haplotypes and phase uncertainty adversely affect statistical power. Multilocus "scoring" approaches that capitalize on marginal effects of individual SNPs are being developed as well. These approaches indirectly take into account the interaction between SNPs while adjusting for LD[26,40]. In particular, these approaches are expected to have good power under models inducing substantial marginal SNP effects and strong LD. Although haplotype-based approaches and scoring methods such as the generalized T^2 test provide relatively high power in the respective situations, it has been noted that the extent of LD can be markedly different between the case and the control groups in a region of genetic association[8]. Therefore, a case-control LD comparison appears to be a promising addition to existing methods of characterizing multilocus associations. Further, Nielsen *et*

al.[10] examined a two-SNP situation and found that a test comparing LD coefficients can be more powerful than a single-locus or a haplotypic test.

We extend these results to the case of multiple SNPs. The LD contrast test, like any other method, would not be expected to have superior power across all susceptibility models. One virtue of the LD contrast test is the reduced number of parameters (e.g. there is only a single LD coefficient in two-SNP case, although there are four haplotypes), and as we discuss here, there are certain models that may result in good power of the LD contrast test. A prominent example is when multilocus susceptibilities induce "orthogonal"-like distribution of dilocus haplotypes in cases and controls. Heterogeneity models where mutations are associated with haplotypes that are distinct with respect to a large proportion of the alleles that they carry, may result in such orthogonality for some of the dilocus pairs. To illustrate why the LD contrast test can work well under these scenarios, denote two alleles at either of two loci by 1 and 2: $A \equiv 1, B \equiv 1, a \equiv 2, b \equiv 2$. The disequilibrium coefficient can be written in terms of the haplotype frequencies as $D_{AB} \equiv D_{11} = P_{11}P_{22} - P_{12}P_{21}$. The disequilibrium is large in a particular sample if the haplotypes "11", "22" are over-represented compared to the two other types. Therefore the ratio $D_{AB}^{\text{cases}}/D_{AB}^{\text{controls}}$ tends to be away from one when the "orthogonal" haplotypes "11" and "22" are over-represented in one of the groups. A similar situation holds with the composite LD definition, because the value of the sum $D_{AB} + D_{A/B}$ increases with D_{AB}. Moreover, we found that while the T^2 test and the LD contrast test provide similar power under a general diplotype-driven model, the correlation between p-values of the two tests is

low (the left graph of Figure 3), suggesting that these approaches distinguish between different aspects of association. That is, the LD comparison tests are more sensitive to interactions that tend to induce small marginal effects associated with individual SNPs. The magnitude of marginal effects is largely unpredictable in practice as it is determined by both multilocus susceptibility values and the corresponding population frequencies. Therefore, it would vary from population to population given the same penetrance configuration.

The question remains, which measure, Δ' or the correlation r is more appropriate for the comparison of patterns of LD. Two samples can have equal LD correlation values, while the standardized LD coefficients are unequal, and vice versa[41], making the *a priori* choice of the statistic somewhat difficult. An appealing feature of the standardization by the LD bounds ("di-prime-ization") in that it makes the measure independent of single locus frequencies. The independence is in the sense of the range that the coefficient can take. The allele frequencies very much remain a part of that definition[42,43], and it would be a mistake to interpret the standardized coefficient as being free of dependencies on the allele or genotype frequencies. On the other hand, the correlation coefficient enjoys a well-defined statistical and population-genetic properties and gives a straightforward extension to the principal components based inference. The simulations (II) show somewhat higher power of the tests based on the correlation, and the CYP2D6 data set considered here provides an example where the test based on the correlation provides a slightly stronger evidence of association ($p = 0.033$ vs. $p = 0.061$), although both results

can be considered indicative of association. In the absence of a specific hypothesis it seems

reasonable to employ the correlation-based analysis, and reserve the Δ'-based LD

comparisons for more detailed characterizations of LD. Nevertheless, the correlation and

the the Δ'-based comparisons address different hypotheses, and both tests have their value.

Our choice of a particular simulation design might have favored greater deviations from the

equality of correlations. The orthogonal model ("Simulation I") in which the correlation

and the Δ' values were almost identical, showed similar power of two tests.

In our simulations we found that the squared difference-based statistic has better

power than the statistic based on the comparison of the principal components, although

the correlation between p-values obtained for these tests is high. The squared difference is

a more omnibus test. For example, if the amount of LD is proportionally higher among the

cases for all pairs of markers, the principal components test will lack power. In addition,

there is uncertainty in the number of components to use. Still, such tests can provide a

description of the multivariate structure of LD. The principal component based analysis

seems to be most valuable at the descriptive stage once the association is established. The

default value for the number of principal components (k) can be set to Krzanowski's

recommendation[25] to use the largest integer k smaller than $L/2$.

In summary, we suggest that statistical approaches to compare pairwise LD matrices

between the case and the control samples are useful additions to already available statistical

mapping tools. As with single marker case-control analysis, population heterogeneity is an

issue. Further research should emphasize extending these methods to accommodate family

data and provide methods robust to population stratification and admixture.

Acknowledgments

References

1. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 70: 425–434

2. Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG (2002) Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 53: 79–91

3. Stram DO, Leigh Pearce C, Bretsky P, Freedman M, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Thomas DC (2003) Modeling and E-M estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals. Hum Hered 55: 179–190

4. Epstein MP, Satten GA (2003) Inference on haplotype effects in case-control studies using unphased genotype data. Am J Hum Genet 73: 1316–1329

5. Lin DY. (2004) Haplotype-based association analysis in cohort studies of unrelated individuals. Genet. Epidemiol 26: 255–264

6. Shibata K, Ito T, Kitamura Y, Iwasaki N, Tanaka H, Kamatani N (2004) Simultaneous

estimation of haplotype frequencies and quantitative trait parameters: applications to the test of association between phenotype and diplotype configuration. Genetics 168: 525–539

7. Morris AP (2005) Direct analysis of unphased SNP genotype data in population-based association studies via Bayesian partition modelling of haplotypes. Genet Epidemiol 29: 91–107

8. Hayes MG, Roe CA, Ng M, Bosque-Plata L, T. Tsuchiya T, Wu X, Ambrose NG, Yairi E, Cook EH, Cox NJ (2004) Case-control differences in linkage disequilibrium as a tool for gene mapping in complex diseases. Am J Hum Genet (suppl; Proceedings of the Am Soc Hum Genet Annual Meeting)

9. Abecasis GR, Cookson WO (2000) GOLD–graphical overview of linkage disequilibrium. Bioinformatics 16: 182–183

10. Nielsen DM, Ehm MG, Zaykin DV, Weir BS (2004) Effect of two- and three-locus linkage disequilibrium on the power to detect marker/phenotype associations. Genetics 168: 1029–1040

11. Nielsen DM, Ehm MG, Weir BS (1999) Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. Am J Human Genet 63: 1531–1540

12. Wittke-Thompson JK, Pluzhnikov A, Cox NJ (2005) Rational inferences about

departures from Hardy-Weinberg equilibrium. Am J Hum Genet 76: 967–986

13. Schaid DJ (2004a) Linkage disequilibrium testing when linkage phase is unknown. Genetics 166: 505–512

14. Zaykin DV (2004) Bounds and normalization of the composite linkage disequilibrium coefficient. Genet Epidemiol 27: 252–257

15. Hamilton DC, Cole DE (2004) Standardizing a composite measure of linkage disequilibrium. Ann Hum Genet 68: 234–239

16. Weir BS (1996) Genetic Data Analysis II. Sinauer Associates, Sunderland, MA

17. Huttley GA, Wilson SR (2000) Testing for concordant equilibria between population samples. Genetics 156: 2127–2135

18. Lewontin RC (1964) The interaction of selection and linkage. I. general considerations; heterotic models. Genetics 49: 49–67

19. Yule GU (1912) On the methods of measuring association between two attributes. J Royal Stat Soc 75: 579–642

20. Weir BS (1979) Inferences about linkage disequilibrium. Biometrics 35: 235–254

21. Peduzzi PN, Detre KM, Chan YK (1983) Upper and lower bounds for correlations in 2 × 2 tables – revisited. J Chronic Dis 36: 491–496

22. Weir BS, Cockerham CC (1979) Estimation of linkage disequilibrium in randomly mating populations. Heredity 42: 105–111

23. Meng Z, Zaykin DV, Xu CF, Wagner M, Ehm MG (2003) Selection of genetic markers for association analyses, using linkage disequilibrium and haplotypes. Am J Hum Genet 73: 115–130

24. Krzanowski WJ (1979) Between-groups comparison of principal components. J Am Stat Soc 74: 703–707

25. Krzanowski WJ (1993) Permutational tests for correlation matrices. Statistics and Computing 3: 37–44

26. Xiong M, Zhao J, Boerwinkle E (2002) Generalized T^2 test for genome association studies. Am J Hum Genet 70: 1257–1268

27. Westfall PH, Zaykin DV, Young SS (2002) Multiple tests for genetic effects in association studies. Methods Mol Biol 184: 143–168

28. Evans WE, Relling MV (1999) Pharmacogenomics: translating functional genomics

into rational therapeutics. Science 286: 487–491

29. Hosking LK, Boyd PR, Xu CF, Nissum M, Cantone K, Purvis IJ, Khakhar R, Barnes MR, Liberwirth U, Hagen-Mann K, Ehm MG, Riley JH (2002) Linkage disequilibrium mapping identifies a 390 kb region associated with CYP2D6 poor drug metabolising activity. Pharmacogenomics J 2: 165–175

30. Van Buuren S, Oudshoorn CGM (2000) Multivariate Imputation by Chained Equations: MICE V1.0 User's manual. Report PG/VGZ/00.038, TNO Prevention and Health, Leiden

31. Murdoch DJ, Chow ED (1996) A graphical display of large correlation matrices. The American Statistician 50: 178–180

32. Culverhouse R, Klein T, Shannon W (2004) Detecting epistatic interactions contributing to quantitative traits. Genet Epidemiol 27: 141–152

33. Almasy L, Terwilliger JD, Nielsen D, Dyer TD, Zaykin D, Blangero J (2001) GAW12: simulated genome scan, sequence, and family data for a common disease. Genet Epidemiol 21: (S) 332–338

34. Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138: 963–971

35. Rubio JP, Bahlo M, Butzkueven H, van Der Mei IA, Sale MM, Dickinson JL, Groom P, Johnson LJ, Simmons RD, Tait B, Varney M, Taylor B, Dwyer T, Williamson R, Gough NM, Kilpatrick TJ, Speed TP, Foote SJ (2002) Genetic dissection of the human leukocyte antigen region by use of haplotypes of Tasmanians with multiple sclerosis. Am J Hum Genet 70: 1125–1137

36. Suarez BK, Parsian A, Hampe CL, Todd RD, Reich T, Cloninger CR. (1994) Linkage disequilibria at the D2 dopamine receptor locus (DRD2) in alcoholics and controls. Genomics. 9: 12–20.

37. Box GEP (1957) Non-Normality and Tests on Variances. Biometrika 40: 318–335

38. Boos DD, Brownie C (2004) Comparing variances and other measures of dispersion. Statistical Science 19: 571–578

39. Schaid DJ (2004b) Evaluating associations of haplotypes with traits. Genet Epidemiol 27: 348–364

40. Clayton D, Chapman J, Cooper J (2004) Use of unphased multilocus genotype data in indirect association studies. Genet Epidemiol 27: 415–428

41. Zapata C, Alvarez G (1997) Testing for homogeneity of gametic disequilibrium among populations. Evolution 51: 606–607

42. Hedrick PH (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117: 331–341

43. Lewontin RC (1988) On measures of gametic disequilibrium. Genetics 120: 849–852

Table 1. Population case-control $\hat{\Delta}'_{AB}$ matrix for the 6-SNP haplotype heterogeneity model (Simulation I)

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
SNP1		0.127	0.192	0.193	0.128	-0.181
SNP2	0.055		0.185	0.104	0.236	-0.048
SNP3	0.143	0.089		0.185	-0.138	0.268
SNP4	0.077	0.005	0.133		0.193	0.055
SNP5	0.042	0.187	-0.208	0.085		-0.085
SNP6	-0.229	-0.108	0.160	0.045	-0.118	

Upper part of the matrix: $\hat{\Delta}'_{AB}$ among the cases.

Lower part of the matrix: $\hat{\Delta}'_{AB}$ among the controls.

Table 2. Power values for the 6-SNP haplotype heterogeneity model (Simulation I)

Prevalence	Haplotype Test	T^2	Z_1 $(k{=}1)$	Z_1 $(k{=}2)$	Z_2 (corr)	Z_2 (Δ')	min$P$
0.05	0.388	0.081	0.365	0.417	0.663	0.654	0.073
0.50	0.251	0.081	0.196	0.301	0.454	0.443	0.073

$k = 1$ for Z_1 test corresponds to about 15% of variation accounted by principal components.

$k = 2$ for Z_1 test corresponds to about 25% of variation accounted by principal components.

min$P$: single-SNP permutation-based trend (allelic) test.

Table 3. Power values for the diplotype-driven model with 15 and 30 SNPs (Simulation II)

SNPs	T^2	Z_1 (50% var)	Z_1 (75% var)	Z_1 (90% var)	Z_2 (corr)	Z_2 (Δ')	min$P$
15	0.707	0.478	0.423	0.407	0.711	0.617	0.569
30	0.818	0.563	0.621	0.488	0.847	0.773	0.171

min$P$: single-SNP permutation-based trend (allelic) test.

Figure legends

Figure 1: LD ellipse plots for the CYP2D6 data. Left graph: $\hat{\mathbf{r}}_Y$ vs. $\hat{\mathbf{r}}_N$. Right graph: $\hat{\boldsymbol{\Delta}}'_Y$ vs. $\hat{\boldsymbol{\Delta}}'_N$.

Figure 2: Composite LD color plot (sample dataset from Simulation II with 29 SNPs). Above the diagonal: $\hat{\boldsymbol{\Delta}}'_Y$ below the diagonal: $\hat{\boldsymbol{\Delta}}'_N$. Δ'-based LD difference p-value: $< 1 \times 10^{-3}$. Blue to red corresponds to the increase in abs(Δ').

Figure 3: Correspondence between p-values (simulation II with 15 SNPs). Left graph: plot of T^2 vs. Z_2 -log of p-values (low correlation). Right graph: plot of Z_1 vs. Z_2 -log of p-values (high correlation).
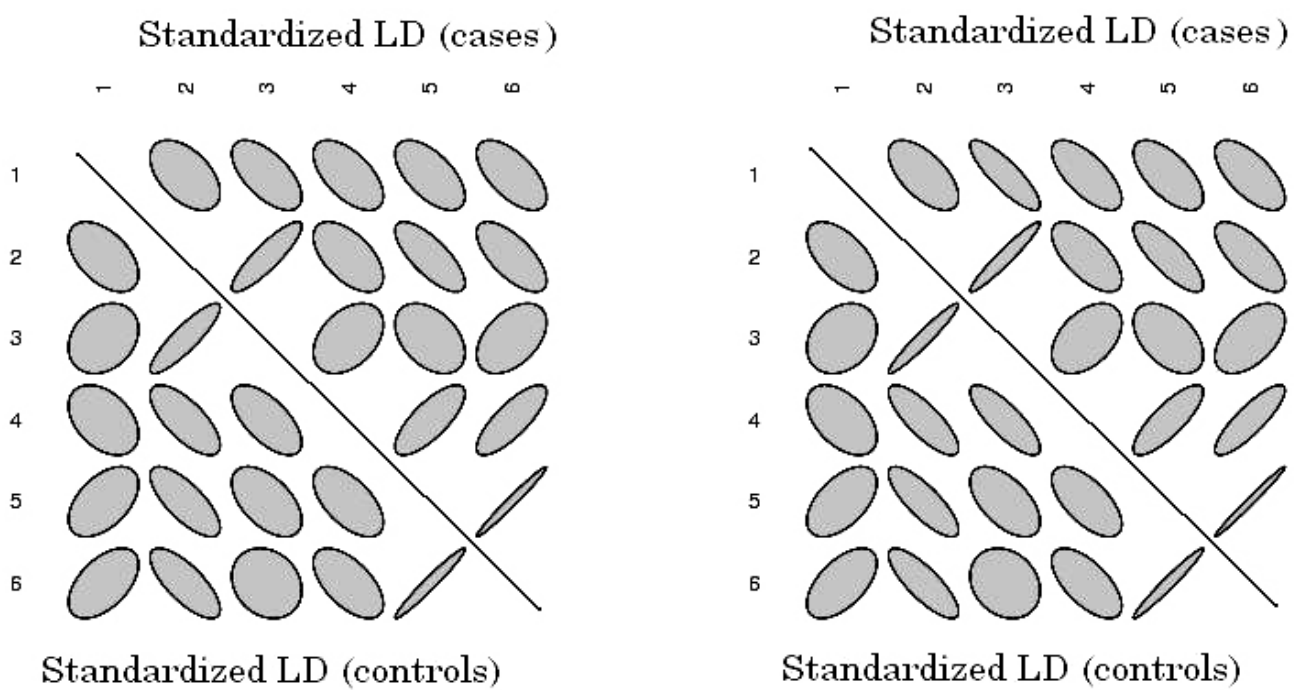
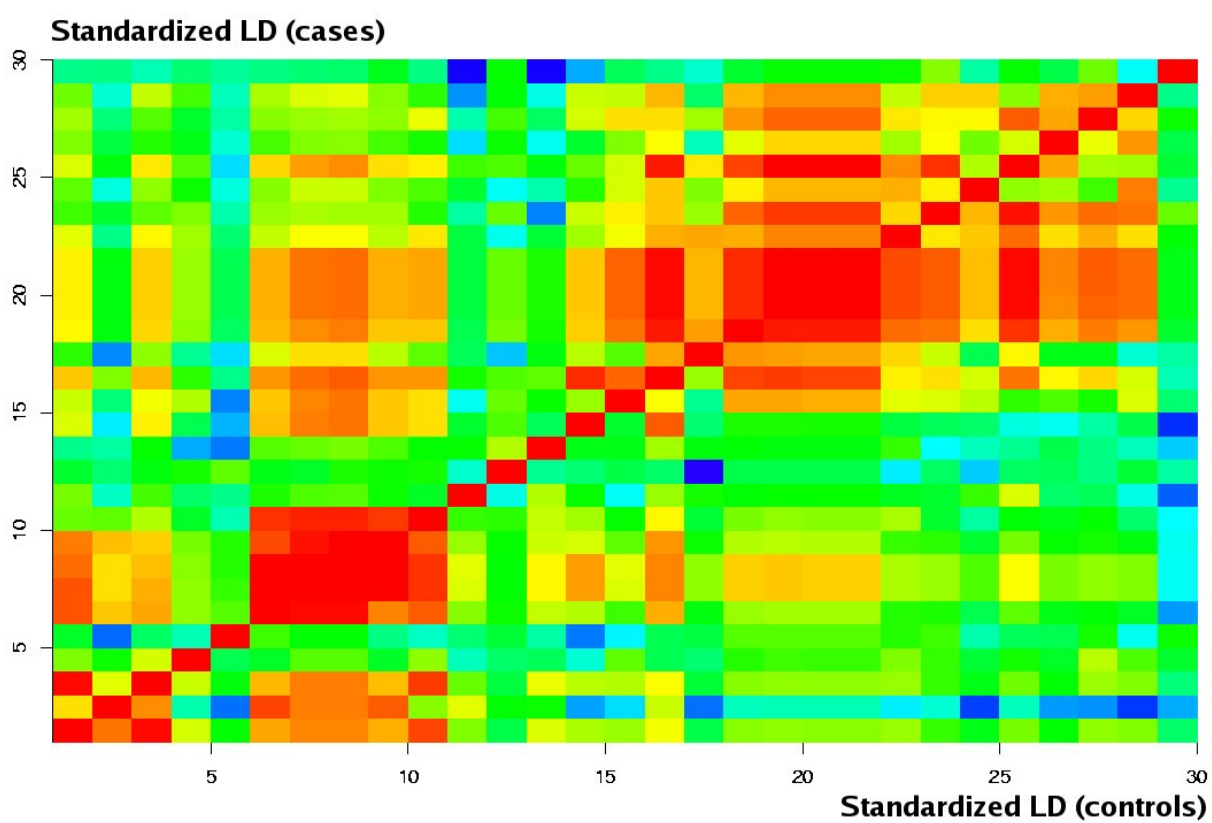Figure 1: (description is on the Figure Legends page)
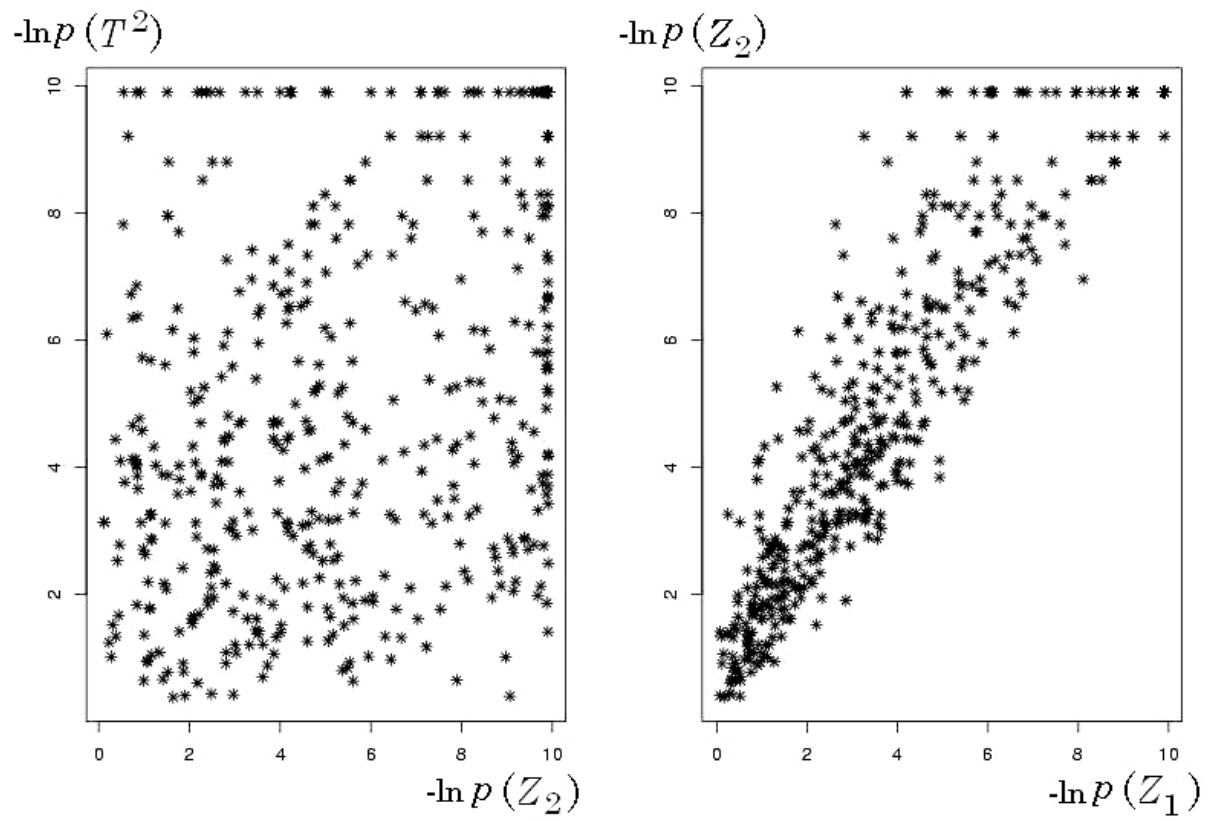
Figure 2: (description is on the Figure Legends page)

Figure 3: (description is on the Figure Legends page)