Haplotype associations with quantitative traits in the presence of complex multilocus and heterogeneous effects

Kyoko Shibata[1], **Luda Diatchenko**[2], and **Dmitri V. Zaykin**[1],[†]

1 *National Institute of Environmental Health Sciences, National Institutes of Health*

2 *Center for Neurosensory Disorders; University of North Carolina at Chapel Hill*

Abstract

In genetic mapping of complex traits, scored haplotypes are likely to represent only a subset of all causal polymorphisms. At the extreme of this scenario, observed polymorphisms are not themselves functional, and only linked to causal ones via linkage disequilibrium (LD). We will demonstrate that due to such incomplete knowledge regarding the underlying genetic mechanism, the variance of a trait may become different between the scored haplotypes. Thus, unequal variances between haplotypes may be indicative of additional functional polymorphisms affecting the trait. Methods accounting for such haplotype-specific variance may also provide an increased power to detect complex associations. We suggest ways to estimate and test these haplotypic variance contrasts, and incorporate them into the haplotypic tests for association. We further extend this approach to data with unknown gametic phase via likelihood-based simultaneous estimation of haplotypic effects and their frequencies. We find our approach to provide additional power, especially under the following types of models: (a) where scored and unobserved variants are epistatically interacting with each other; and (b) under heterogeneity models, where multiple unobserved mutations are linked to nonfunctional observed polymorphisms via LD. An illustrative example of usefulness of the method is discussed, utilizing analysis of multilocus effects within the catechol-O-methyl transferase (COMT) gene.

Keywords

haplotype association; complex disease; epistasis; gene-gene interactions

Introduction

There is a substantial body of work on characterizing haplotype-trait associations using unrelated individuals. Simple methods of dealing with the unknown haplotype phase in association tests have been proposed [Schaid et al., 2002; Zaykin et al., 2002; Luo et al., 2006]. Xie and Stram [2005] showed asymptotic equivalence of two common types of these approaches. These methods have been found to provide adequate inference concerning both hypothesis testing, and association parameter estimation, and have been recommended for usage [Stram et al., 2003; Kraft et al., 2005; Xie and Stram, 2005; Kraft and Stram, 2007]. Tzeng et al. [2006] extended these approaches to incorporate evolutionary clustering of haplotypes. Unbiased estimation of association parameters may require a simultaneous estimation of haplotype frequencies and association parameters. The maximum-likelihood

†Corresponding author: Dmitri Zaykin, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, South Bldg (101)/B356B, Mail Drop: A3-03, P.O. BOX 12233, RTP, NC 27709, fax: (919) 541-4311, phone: (919) 541-0096, email: E-mail: zaykind@niehs.nih.gov.

methods are theoretically preferable [Lin and Huang, 2007; Allen and Satten, 2008]. A variety of such methods have been proposed [Tregouet et al., 2002; Epstein and Satten, 2003; Stram et al., 2003; Zhao et al., 2003; Shibata et al., 2004; Lin et al., 2005; Lin and Zeng, 2006]. In this article, we incorporate haplotype and diplotype-specific variances into the likelihood for unphased data. The motivation for this extension comes from a scenario where haplotypes under investigation are either linked via LD to causal variants, or represent only a part of all causal variation. In both cases, the effect of an observed variant (i.e. haplotype) on the trait (Y) is a weighted average of the effects that correspond to all relevant polymorphisms considered jointly. The weights are given by the frequencies of these unobserved joint polymorphisms. To simplify the exposition, we first assume that the observed variant is an SNP "A" with alleles A_1, A_2. We denote haplotypes that include A_1 by h_{A_1}, and assume no dominance effects. Then the effect of A_1, i.e. the expected value $E(Y|A_1)$ is

$$\mu_{A_1} = \frac{\sum_{j \in h_{A_1}} \mu_j \, p_j}{\sum_{j \in h_{A_1}} p_j}$$

(1)

where j is indexing over all haplotypes h_{A_1} that carry the allele A_1, μ_j denotes $E(Y|h_j)$, and p_j is the population frequency of the j-th haplotype (h_j). Unless A is the only polymorphism that is relevant with regard to the trait under study, there is uncertainty associated with the frequencies p_j, because the haplotypes are partially typed. This leads to unpredictability of the effect size, μ_{A_1}. As the frequencies vary, the observed effect changes in value, and might even change its sign [Lin et al., 2007; Zaykin and Shibata, 2008], thus the power to detect associations can be greatly reduced. This is a consequence of "partial knowledge": if the causative haplotypes are completely typed, the effects are just μ_j, independent of the frequencies.

Using a similar notation, the variance of a trait among individuals carrying A_1 is

$$V_{A_1} = \frac{\sum_{j \in h_{A_1}} p_j V_j}{\sum_{j \in h_{A_1}} p_j} + \frac{\sum_{j \in h_{A_1}} p_j (\mu_j - \mu_{A_1})^2}{\sum_{j \in h_{A_1}} p_j}$$

(2)

where $V_{A_1} \equiv \mathrm{Var}(Y|A_1)$, and $V_j \equiv \mathrm{Var}(Y|h_j)$ - the variance among individuals carrying a haplotype h_j. We can assume that all V_j are equal to a common value, σ^2, and even then the allele-specific variance remains elevated compared to σ^2, due to differences between the underlying effects:

$$V_{A_1} = \sigma^2 + \frac{\sum_{j \in h_{A_1}} p_j (\mu_j - \mu_{A_1})^2}{\sum_{j \in h_{A_1}} p_j}$$

(3)

The contrast between the variances, V_{A_1} vs. V_{A_2} can also be unequal, and this can be incorporated into association methods. Models where A is either a functional locus, or a variant in LD with nearby causative polymorphisms can produce situations where $V_{A_1} \neq V_{A_2}$, even when the population frequencies p_j are such that there is no effect at the observed locus, $\mu_{A_1} = \mu_{A_2}$ [Zaykin and Shibata, 2008].

The main idea of the approach that we develop here is to estimate and test the variance contrast, V_{A_1} vs. V_{A_2}, in addition to, or simultaneously with the usual inference about equality of the genetic effects, μ_{A_1} vs. μ_{A_2}. We extend this approach to the case where A is represented by multi-SNP haplotypes, estimated from unphased, single-locus genotype data.

Our approach is based on random population samples, and we assume that there is no statistical confounding, such as confounding due to population stratification. In the absence of confounding, both mean and the variance contrasts are interpreted in a similar manner: $\mu_{A_1} \neq \mu_{A_2}$, or $V_{A_1} \neq V_{A_2}$ indicate either (a) a functional involvement of the observed polymorphisms, or (b) association via LD with nearby, unobserved functional variants [Zaykin and Shibata, 2008]. In the first case, the unobserved factor, B (e.g. a factor with two levels, B_1, B_2) can be non-genetic, such as an environmental factor that interacts with the observed variation.

Methods

We are concerned with haplotype and diplotype association methods that are capable of dealing with unphased data. Nevertheless, we start with a single di-allelic (A_1, A_2) SNP, because this allows us to succinctly describe the essence of the methods. We define the following three null hypotheses (H_0) of interest:

$$H_0^{\mu,V}:(\mu_{A_1}=\mu_{A_2};V_{A_1}=V_{A_2})$$
$$H_0^{\mu}:(\mu_{A_1}=\mu_{A_2})$$
$$H_0^{V}:(V_{A_1}=V_{A_2})$$

(4)

with the corresponding likelihood ratio test statistics. A rejection of the more general hypothesis, $H_0^{\mu,V}$, indicates that there are differences in either means or the variances of the trait between the alleles A_1 and A_2. The other two hypotheses are formed specifically regarding the allelic means or the variances. Genotypic rather than allelic-based hypotheses and tests are similarly defined. For example, the H_0 regarding the genotypic means is stated as H_0^{μ}: ($\mu_{A_1A_1} = \mu_{A_1A_2} = \mu_{A_2A_2}$). The likelihoods are constructed assuming the normal distribution for the trait. Log-likelihood contributions are added over individuals in the sample, which are assumed to be independent. L_0 in the following subsections refers to the log-likelihood under the common mean and the variance for all alleles or genotypes; L_1 allows for allele- or genotype-specific means and variances, L_2 allows for allele- or genotype-specific means and common variance, and L_3 allows for allele-or genotype-specific variances. The estimates $\hat{\mu}$, $\hat{V}$ used in the likelihoods are specific to alleles or genotypes at the scored locus, A: it appears difficult to take advantage of incorporating mixture proportions p_j given in equations (1, 2, 3), since these depend on the unknown factor, B.

Allelic-based tests

The model that corresponds to allelic tests and estimates is also referred to as the additive model. To construct allelic likelihood ratio test statistics (LRTs) that correspond to our hypotheses, $H_0^{\mu,V}, H_0^{\mu}, H_0^{V}$, we define the following log-likelihoods:

$$L_0=2\sum_{i=1}^{n}\ln\varphi(Y_i;\widehat{\mu},\widehat{V})$$
$$L_1=L_1^{A_1A_1}+L_1^{A_1A_2}+L_1^{A_2A_2}$$
$$L_2=L_2^{A_1A_1}+L_2^{A_1A_2}+L_2^{A_2A_2}$$
$$L_3=L_3^{A_1A_1}+L_3^{A_1A_2}+L_3^{A_2A_2}$$

(5)

where

$$L_1^{A_1A_1}=2\sum_{i\in A_1A_1}^{n_{A_1A_1}}\ln\varphi(Y_i;\widehat{\mu}_{A_1},\widehat{V}_{A_1}); \quad L_1^{A_2A_2}=2\sum_{i\in A_2A_2}^{n_{A_2A_2}}\ln\varphi(Y_i;\widehat{\mu}_{A_2},\widehat{V}_{A_2})$$

$$L_1^{A_1A_2}=\sum_{i\in A_1A_2}^{n_{A_1A_2}}\ln\varphi(Y_i;\widehat{\mu}_{A_1},\widehat{V}_{A_1})+\sum_{i\in A_1A_2}^{n_{A_1A_2}}\ln\varphi(Y_i;\widehat{\mu}_{A_2},\widehat{V}_{A_2})$$

$$L_2^{A_1A_1}=2\sum_{i\in A_1A_1}^{n_{A_1A_1}}\ln\varphi(Y_i;\widehat{\mu}_{A_1},\widehat{V}); \quad L_2^{A_2A_2}=2\sum_{i\in A_2A_2}^{n_{A_2A_2}}\ln\varphi(Y_i;\widehat{\mu}_{A_2},\widehat{V})$$

$$L_2^{A_1A_2}=\sum_{i\in A_1A_2}^{n_{A_1A_2}}\ln\varphi(Y_i;\widehat{\mu}_{A_1},\widehat{V})+\sum_{i\in A_1A_2}^{n_{A_1A_2}}\ln\varphi(Y_i;\widehat{\mu}_{A_2},\widehat{V})$$

$$L_3^{A_1A_1}=2\sum_{i\in A_1A_1}^{n_{A_1A_1}}\ln\varphi(Y_i;\widehat{\mu},\widehat{V}_{A_1}); \quad L_3^{A_2A_2}=2\sum_{i\in A_2A_2}^{n_{A_2A_2}}\ln\varphi(Y_i;\widehat{\mu},\widehat{V}_{A_2})$$

$$L_3^{A_1A_2}=\sum_{i\in A_1A_2}^{n_{A_1A_2}}\ln\varphi(Y_i;\widehat{\mu},\widehat{V}_{A_1})+\sum_{i\in A_1A_2}^{n_{A_1A_2}}\ln\varphi(Y_i;\widehat{\mu},\widehat{V}_{A_2})$$

$$\widehat{\mu}_{A_1}=\frac{2\sum_{i\in A_1A_1}^{n_{A_1A_1}}Y_i+\sum_{i\in A_1A_2}^{n_{A_1A_2}}Y_i}{2n_{A_1A_1}+n_{A_1A_2}}; \quad \widehat{\mu}_{A_2}=\frac{2\sum_{i\in A_2A_2}^{n_{A_2A_2}}Y_i+\sum_{i\in A_1A_2}^{n_{A_1A_2}}Y_i}{2n_{A_2A_2}+n_{A_1A_2}}$$

$$\widehat{V}_{A_1}=\frac{2\sum_{i\in A_1A_1}^{n_{A_1A_1}}(Y_i-\widehat{\mu}_{A_1})^2+\sum_{i\in A_1A_2}^{n_{A_1A_2}}(Y_i-\widehat{\mu}_{A_1})^2}{2n_{A_1A_1}+n_{A_1A_2}}$$

$$\widehat{V}_{A_2}=\frac{2\sum_{i\in A_2A_2}^{n_{A_2A_2}}(Y_i-\widehat{\mu}_{A_2})^2+\sum_{i\in A_1A_2}^{n_{A_1A_2}}(Y_i-\widehat{\mu}_{A_2})^2}{2n_{A_2A_2}+n_{A_1A_2}}$$

$$\widehat{V}=\frac{2\sum_{i\in A_1A_1}^{n_{A_1A_1}}(Y_i-\widehat{\mu}_{A_1})^2+\sum_{i\in A_1A_2}^{n_{A_1A_2}}(Y_i-\widehat{\mu}_{A_1})^2}{2n}$$
$$+\frac{2\sum_{i\in A_2A_2}^{n_{A_2A_2}}(Y_i-\widehat{\mu}_{A_2})^2+\sum_{i\in A_1A_2}^{n_{A_1A_2}}(Y_i-\widehat{\mu}_{A_2})^2}{2n}$$

$$\widehat{\mu}=\frac{\sum_{i=1}^{n}Y_i}{n}; \quad \widehat{V}_0=\frac{\sum_{i=1}^{n}(Y_i-\widehat{\mu})^2}{n}$$

(6)

where $\varphi(y; \mu, V)$ is the normal density, $N(\mu, V)$, evaluated at y, and $n_{A_1A_1}$, $n_{A_1A_2}$, $n_{A_2A_2}$ are the A_1A_1, A_1A_2, A_2A_2 genotype counts, $n_{A_1A_1} + n_{A_1A_2} + n_{A_2A_2} = n$.

Tests for the means as well as tests for the variances can be constructed while allowing for the other parameter to be different between alleles. There is a concern that the test for the mean with unequal variances (which we denote by T^{μ^*}) might be less robust than the test based on the common variance, T^μ. Thus, we investigate both versions of the LRTs. The LRT for the variance is denoted by T^V. The combined test for the both parameters, $T^{\mu,V}$, is obtained in a straightforward way as a difference of the likelihoods $2(L_1 - L_0)$. This difference is equal to the sum, $T^\mu + T^V$. We also consider a different combined test, $T^{\mu^*,V}$, based on the sum, $T^{\mu^*} + T^V$. The LRTs are:

$$T^{\mu,V}=2(L_1 - L_0) \sim \chi^2_{(2)}$$
$$T^{\mu^*,V}=2(2L_1 - L_2 - L_3) \sim \chi^2_{(2)}$$
$$T^\mu=2(L_2 - L_0) \sim \chi^2_{(1)}$$
$$T^{\mu^*}=2(L_1 - L_3) \sim \chi^2_{(1)}$$
$$T^V=2(L_1 - L_2) \sim \chi^2_{(1)}$$

(7)

The notation indicates that, for example, $T^{\mu,V}$, has an asymptotic chi-square distribution with two degrees of freedom. The degrees of freedom are given by the difference in the number of parameters between the likelihoods forming the LRT.

Genotypic-based tests

Genotypic rather than allelic-based tests are constructed similarly. The mean and the variance estimators for the homozygote A_1A_1 are obtained as

$$\widehat{\mu}_{A_1A_1} = \frac{\sum_{i \in A_1A_1}^{n_{A_1A_1}} Y_i}{n_{A_1A_1}}; \ \widehat{V}_{A_1A_1} = \frac{\sum_{i \in A_1A_1}^{n_{A_1A_1}} (Y_i - \widehat{\mu}_{A_1A_1})^2}{n_{A_1A_1}}$$

(8)

These are similarly defined for the genotypes A_1A_2 and A_2A_2. The log-likelihoods are

$$L_0 = \sum_{i=1}^{n} \ln\varphi(Y_i; \widehat{\mu}, \widehat{V})$$

$$L_1 = \sum_{i \in A_1A_1}^{n_{A_1A_1}} \ln\varphi(Y_i; \widehat{\mu}_{A_1A_1}, \widehat{V}_{A_1A_1}) + \sum_{i \in A_1A_2}^{n_{A_1A_2}} \ln\varphi(Y_i; \widehat{\mu}_{A_1A_2}, \widehat{V}_{A_1A_2})$$

$$+ \sum_{i \in A_2A_2}^{n_{A_2A_2}} \ln\varphi(Y_i; \widehat{\mu}_{A_2A_2}, \widehat{V}_{A_2A_2})$$

$$L_2 = \sum_{i \in A_1A_1}^{n_{A_1A_1}} \ln\varphi(Y_i; \widehat{\mu}_{A_1A_1}, \widehat{V}) + \sum_{i \in A_1A_2}^{n_{A_1A_2}} \ln\varphi(Y_i; \widehat{\mu}_{A_1A_2}, \widehat{V})$$

$$+ \sum_{i \in A_2A_2}^{n_{A_2A_2}} \ln\varphi(Y_i; \widehat{\mu}_{A_2A_2}, \widehat{V})$$

$$L_3 = \sum_{i \in A_1A_1}^{n_{A_1A_1}} \ln\varphi(Y_i; \widehat{\mu}, \widehat{V}_{A_1A_1}) + \sum_{i \in A_1A_2}^{n_{A_1A_2}} \ln\varphi(Y_i; \widehat{\mu}, \widehat{V}_{A_1A_2})$$

$$+ \sum_{i \in A_2A_2}^{n_{A_2A_2}} \ln\varphi(Y_i; \widehat{\mu}, \widehat{V}_{A_2A_2})$$

(9)

and the test statistics are

$$T^{\mu,V} = 2(L_1 - L_0) \sim \chi^2_{(4)}$$
$$T^{\mu^*,V} = 2(2L_1 - L_2 - L_3) \sim \chi^2_{(4)}$$
$$T^{\mu} = 2(L_2 - L_0) \sim \chi^2_{(2)}$$
$$T^{\mu^*} = 2(L_1 - L_3) \sim \chi^2_{(2)}$$
$$T^V = 2(L_1 - L_2) \sim \chi^2_{(2)}$$

(10)

The dominant and the recessive models are defined in the same way as the full genotypic, but without making distinctions between certain genotype classes. For example, in the dominant model we estimate $\hat{\mu}_{A1A1 + A1A2}$, $\hat{V}_{A1A1+A1A2}$, $\hat{\mu}_{A2A2}$, $\hat{V}_{A2A2}$, and form the likelihoods according to these four, rather than six parameters.

Haplotype and diplotype tests for unphased data

The LRT statistics for the situation with multiple SNPs and the unknown haplotype phase have a form similar to the statistics just described. However, because of the phase uncertainty, the form of a likelihood function is now

$$L \propto \prod_{i}^{n} \sum_{\text{haplotype pairs}} \Pr(h_s, h_r | \mathbf{H}) \, \varphi(Y_i; \widehat{\mu}_{sr}, \widehat{V}_{sr} | h_r, h_s)$$

(11)

where $\mathbf{H}$ are sample haplotype frequencies and $\Pr(h_s h_r | \mathbf{H})$ is a per-subject probability of a haplotype pair h_s, h_r, under the assumption of Hardy-Weinberg equilibrium (HWE). Haplotype and diplotype means, variances as well as sample frequencies of haplotypes are estimated simultaneously via an EM algorithm. An extension to multiple haplotypes tested all at once in an overall test is straightforward. The likelihoods described above simply incorporate more parameters, specific to means and variances of multiple haplotypes or diplotypes. When the haplotype phase is known, the overall test statistics for the k haplotypic means can be shown to take the form

$$T^{\mu} = \frac{\sum_{i=1}^{k} n_i (\widehat{\mu}_i - \widehat{\mu})^2}{\widehat{V}}$$

(12)

$$T^{\mu^*} = \sum_{i=1}^{k} \frac{n_i (\widehat{\mu}_i - \widehat{\mu})^2}{\widehat{V}_i},$$

(13)

where n_i are haplotype counts. Under the normal model, the haplotypic overall tests for the means are analogous to the "2N" ANOVA described in Zaykin et al. [2002], where asymptotic comparisons were made with a corresponding trend test. The variance test has the form

$$T^V = \ln(\widehat{V}) \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} n_i \ln(\widehat{V}_i),$$

(14)

This statistic is equivalent to the Bartlett test [Bartlett, 1937]. Under $H_0^{\mu, V}$ and normality, it is independent of the statistic that is based on the ratio of the mean difference and the pooled variance. Thus, asymptotically equivalent forms of the LRTs for testing the means and the LRTs for the variances are asymptotically independent. Frequencies of rare classes become critical with regard to the properties of an overall test. In our comparisons, we used an alternative overall test, the Simes test, based on p-values obtained from tests on individual haplotypes. Simes [1986] described this combination test that allows for dependencies among p-values. The Simes test rejects the overall H_0 at the level α if $p_{(i)} \leq i\alpha/k$ for at least one i, where k is the number of tested haplotypes, and $p_{(i)}$ are ordered p-values. Equivalently, the overall p-value is given by $\min\{kp_{(i)}/i\}$. Thus, this approach is to take one haplotype (h_i) at a time, contrasted against the group of all other haplotypes, $h_{\bar{i}}$. As we will verify by simulations, this test can be more powerful than a multi-parameter overall test when only a small proportion of all haplotypes carry pronounced effects. Although the Simes test is an overall test with regard to the family-wise error rate, it has the same form as the false discovery rate (FDR) controlling procedure by Benjamini and Hochberg [1995] that can be used to make statements for a specific set of haplotypes that clear the FDR threshold.

A note should be made regarding the interpretation of results for the variance-specific tests. When a haplotype h_i is compared against a combined group that consists of all remaining

haplotypes, $h_{\bar{i}}$, a possible heterogeneity of effects combined within $h_{\bar{i}}$ may lead to an increase of the variance associated with that group (Appendix 2). In addition, the variance-specific haplotypic tests are sensitive to a variance contrast that can be induced by strong dominance deviations. Thus, diplotypic, rather than haplotypic tests, especially when the diplotypes h_i/h_i, $h_i/h_{\bar{i}}$ (compared to $h_{\bar{i}}/h_{\bar{i}}$) are found to be associated with an increased variance, provide a more clear evidence for an additional unknown functional variation outside of the studied set of SNPs.

The normality assumption can be relaxed in our implementation by utilizing a resampling distribution of the test statistics for computing p-values of the tests. Asymptotic tests appear to offer a reasonable performance, unless the expected count of a tested variant is low. Nevertheless, it is a good strategy to verify low p-values with a permutational test. Trait transformations to normality, such as the common Box-Cox transformation [Box and Cox, 1964] can also be considered. The common-variance test for the mean, T^μ appears to be robust to deviations from normality (Results section). Similar tests contrasting means for the dominant, recessive, and diplotype models were described in Shibata et al. [2004].

The two combined statistics considered here can be factored as sums, e.g. $T^{\mu,V} = T^\mu + T^V$. That is, these statistics are sums of two chi-square contributions, which are independent under the hypothesis that all means and variances are the same, and assuming normality. At a reviewer's request we will describe the relation of the p-value for this test with p-values obtained via combination methods, from p-values of the two other tests. Denote p-value for testing the means by p_1, and p-value for testing the variances by p_2. Zaykin et al. [2007] described

properties of a method by which L p-values are combined as $p_c = 1 - G_{\lambda L}\left[\sum_{}^{L} G_a^{-1}(1 - p_i)\right]$, where G_x is a cumulative Gamma distribution function with the shape parameter equal to x, and an arbitrary rate parameter. Setting $\lambda = 1$ gives Fisher's combined p-value, which in the case of just two p-values simplifies to $p_c = p_1 p_2 [1 - \ln(p_1 p_2)]$. This p-value corresponds exactly to the p-value for the genotypic/diplotypic test, $T^{\mu,V}$, or $T^{\mu^*,V}$, because Fisher's test is based on

a sum of two degree of freedom chi-squares. The allelic/haplotypic p-value for testing $H_0^{\mu,V}$ is equivalent to p_c obtained with $\lambda = 1/2$, because in this case, single degree of freedom chi-squares are being added. With only two p-values, the combined values are expected to be similar for different values of λ. Thus, the above simplification for Fisher's p_c gives a straightforward way to examine the expected behavior of the test for the combined hypothesis. In particular, the test is expected to provide an additional power only when both of the individual null hypotheses are false. Of note, the solution of $p = p^2[1 - \ln(p^2)]$ shows that when $p_1 = p_2 < \approx 0.285$, the combined value is smaller than the individual values; otherwise, it is larger. Elston [1991] showed that as L becomes large, this "critical value" approaches $1/e$.

Cohort description

To illustrate an application of the proposed association tests, we performed an analysis of the data collected within a larger prospective study, concerned with the pain sensitivity as a risk factor for the development of facial pain [Diatchenko et al., 2005].

Subject recruitment—One hundred ninety six healthy European American pain-free females with an age range of 18 to 34 years were genotyped and phenotyped. Phenotypic procedures and demographic characteristics of the cohort at the time of recruitment were described previously [Diatchenko et al., 2005].

Pain phenotyping procedures—Subjects were phenotyped with respect to their sensitivity to pressure pain, thermal pain, and ischemic pain. A summary measure of thermal threshold was used in this study as it has been shown to be the most sensitive measure with

respect to association with catechol-O-methyltransferase (COMT) genotypes, and has the lowest measurement error [Diatchenko et al., 2006, 2005]. A summary measure was calculated by summing centered and scaled individual measures of thermal pain threshold, collected as subjective estimates of rating thermal stimuli that were applied to the skin overlaying the right masseter muscle, right forearm, and dorsal surface of the right foot. Positive values imply higher pain thresholds, and negative values imply lower pain thresholds.

Genotyping—Genomic DNA was purified using QIAampTM 96 DNA Blood Kit (Qiagen, Valencia, CA, USA) and used for 5′ exonuclease assay [Shi et al., 1999]. The primer and probes were ordered from ABI, Foster City, CA. The genotyping error rate was directly determined and was <0.005. Genotype completion rate was 95%.

Results

We will first present results of theoretical models, followed by the analysis of association of pain characteristics with the COMT polymorphisms. In the theoretical part of the section, we will refer to a model where the observed and the unobserved loci are denoted by A and B, correspondingly. We will study situations where A and B represent loci with multiple haplotypes. When both A and B are either di-allelic, or when A_2, for example, can be considered as the "not-A_1" group of haplotypes, we will denote the vector of four effects, $\mu_{A_iB_j} = E(Y|A_i, B_j)$, by $\mathbf{M}$, and the four population frequencies by $\mathbf{P}$:

Haplotype	$\mathbf{P}$	$\mathbf{M}$
A_1B_1	$p_{A_1B_1}$	$\mu_{A_1B_1}$
A_1B_2	$p_{A_1B_2}$	$\mu_{A_1B_2}$
A_2B_1	$p_{A_2B_1}$	$\mu_{A_2B_1}$
A_2B_2	$p_{A_2B_2}$	$\mu_{A_2B_2}$

Non-interactive models

First, we consider models where both observed (A) and unobserved (B) loci are functional, contributing additively, with no interaction between the loci. In the di-allelic case, the array of effects takes the form $\mathbf{M} = \{\kappa_{A_1} + \kappa_{B_1}, \kappa_{A_1} + \kappa_{B_2}, \kappa_{A_2} + \kappa_{B_1}, \kappa_{A_2} + \kappa_{B_2}\}$, where κ denote additive contributions of respective alleles. Figure 1 shows logs of theoretical mean and variance contrasts for this model: $\ln(\mu_{A_1}/\mu_{A_2})$, and $\ln(V_{A_1}/V_{A_2})$, as functions of the frequency of A_2B_1. The frequency p_{A_1} was set to the frequency of a common COMT haplotype, 0.3034 (Table 1). We used $p_{A_1B_1} = 0.3034$; $p_{A_1B_2} = 0$ in the left graph, and $p_{A_1B_1} = p_{A_1B_2} = 0.3034/2$ in the right graph. We assumed $\kappa_{A_1} = 2$, $\kappa_{A_2} = 3$, $\kappa_{B_1} = 4$, $\kappa_{B_2} = 1$. The graphs show an extended region of $V_{A_1} \neq V_{A_2}$. To evaluate the proposed statistical tests, we took random samples of $n = 500$ from diploid, randomly mating populations, that represent various points along the A_2B_1 frequency axis of the left graph of Figure 1. The phenotype for the subject i with the haplotype pair A_sB_t/A_uB_v in these and the subsequent simulations was generated as $Y_i = \mu A_sB_t + \mu A_uB_v + \xi_i$; where $\xi_i \sim N(0, \sigma^2)$. The locus B was assumed to be unobserved. Unless stated otherwise, all simulations used at least 10,000 samples of 500 individuals and the nominal 5% level for the tests.

The left three columns of Table 2 show resulting power values for a single SNP. Because the goal was to evaluate the *relative* performance of the tests, the σ^2 values, as shown in the table, were taken to be different for each setting, so that medium to high power values could be obtained. The corresponding results for unphased data, with population haplotype frequencies, modeled after the COMT, are given in the right columns of the table. In the case of multiple

unphased haplotypes, the haplotype AGCG corresponds to the allele A_1, while the group A_2 is composed of the seven remaining haplotypes, with their respective frequencies.

The results show a good correspondence of power values for the di-allelic case (Table 2, left three columns), with the case of unphased haplotypes (right columns of Table 2), with some loss of power due to phase ambiguity. There is also a cost of considering diplotype, rather than haplotype associations in this case, because the underlying model consists of phenotypic contributions of haplotypes, and lacks dominance deviations due to entire diplotypes. These findings will be confirmed in the subsequent simulations as well. In these simulations, the unequal variance test for the mean is slightly, but consistently more powerful than the test based on the common variance. This power advantage appears to be confined to cases when the minor allele (or a tested haplotype with frequency less than 0.5) is associated with a decreased variance. There is a power advantage associated with the inclusion of the variance parameter in the range of $p_{A_1B_1}$ from 0.25 to 0.65. On the other hand, the variance contrast in the right graph of Figure 1 gives a substantial power boost only at the left part of the graph (results not shown). The difference between the two graphs is that the frequency of A_1 in the left graph is confined to one of the haplotypes. This results in the complete standardized LD, $D' = 1$. With interactive models, high LD is no longer a requirement for the variance contrast to provide a power advantage, as will be illustrated next.

Interactive models

The previous model was modified to include an epistatic component ε, as $\mathbf{M} = \{\varepsilon + \kappa_{A_1} + \kappa_{B_1}, \kappa_{A_1} + \kappa_{B_2}, \kappa_{A_2} + \kappa_{B_1}, \kappa_{A_2} + \kappa_{B_2}\}$. Figure 2(a) is a such a modification of the model that was used to produce Figure 1(b), now with $\varepsilon = 2$. The effect of it is that the variance contrast curve is pushed upward, compared to the previous picture. This is prominently reflected in the power values for the test that includes the haplotype-specific variance (Table 3). The advantage of including the variance parameter now extends throughout the entire range of the haplotype frequency. Moreover, this finding holds under linkage equilibrium (LE) as well (Figure 2(b), Table 4). Under LE, the behavior of the ratio of the haplotypic means and variances depends only on allele frequencies at the unknown locus [Zaykin and Shibata, 2008], thus the x-axis of Figure 2(b) tracks the frequency of B_1. Power of the the unequal variance test for the mean is slightly worse than that for the test based on the common variance, however, its size is closer to the nominal 5% level in the presence of variance heterogeneity.

Models of association via LD

Models of association via LD (proxy models) assume an unobserved functional locus B, while the observed locus A associated via LD with the B has no functional involvement. In the di-allelic case, the array of effects is $\mathbf{M} = \{x, y, x, y\}$. Graphs for this model (Figure 3) bear similarity to the graphs for the non-interactive model (Figure 1). The distinction is that the behavior in this case is completely governed by LD: both the mean and the variance contrast curves cross zero at the same point. Figure 3(a) shows the LD correlation instead of the D', because D' is equal to one throughout this graph. The power gained from the inclusion of the haplotype-specific variance extends through only a limited part of the graph (Table 5). Nevertheless, in the remainder of the table, the association is so pronounced that it is easily detected with high power by either of the tests ($H_0^{\mu,V}$ or H_0^μ), even though very large values of σ^2 were assumed. For Figure 3(b), where the LD is not complete, there is no advantage in including the haplotype-specific variance into the test (results not shown) - unlike with interactive models, high LD is required for proxy models to induce a substantial variance contrast.

The LD proxy model can be extended to multiple mutations, $B_1, \ldots, B_K$. In this case, the resulting mixture distribution at the observed haplotypes linked via LD with the B may provide

considerable power advantages for the test that includes the variance contrast. The trait distribution at the scored haplotypes is a mixture with weights that depend on p_{B_iA} (Appendix 1). Table 6 lists power values for a model with ten mutations $B_1, \ldots, B_{10}$. The type-I error entries in this case correspond to a situation when we are examining a "wrong" SNP, i.e. an SNP in LE with the mutations B_i. The distribution of the trait in this case is still a mixture, although there is no contrast between the distributions at the scored haplotypes. The inclusion of the variance parameter yields a moderate gain in power. The gain is highest with the allelic/ haplotypic test, because of the lack of dominance deviations attributable to diplotypes. The test T^{μ^*} is found to be slightly more powerful than T^μ in these simulations.

Association of COMT polymorphisms with the pain threshold

Table 7 shows an application of the LRT tests to data on five polymorphisms in the COMT gene. COMT is an example of a gene where multiple interacting variants have been suggested to influence variation in several complex phenotypes. COMT codes for an enzyme that metabolizes catecholamines, such as dopamine (DA), norepinephrine (NE) and epinephrine (Epi), and thus affects an array of cognitive-affective traits, including pain perception [Diatchenko et al., 2005;Nackley et al., 2007;Egan et al., 2001;Enoch et al., 2003;Oroszi and Goldman, 2004;Meyer-Lindenberg et al., 2006]. Common nonsynonymous variation val158met (rs4680) has been shown to influence thermostability of the enzyme [Lotta et al., 1995], however, the associations between the low-activity Met158 allele and numerous complex phenotypes [Egan et al., 2001;Enoch et al., 2003;Karayiorgou et al., 1999;Zubieta et al., 2003;Oroszi and Goldman, 2004] have been often modest, and occasionally inconsistent. Repeated retesting of the val158met has been partially driven by numerous molecular and biochemical studies confirming that Met at position 158 does lead to a 3–4 times lower enzymatic activity of COMT at both cell culture and organismal levels [Lotta et al., 1995;Mannisto and Kaakkola, 1999]. There is evidence that additional functional SNPs in the COMT gene locus can modulate COMT activity, as supported by a number of recent positive association studies and molecular work using cell-based assays [Oroszi and Goldman, 2004;Meyer-Lindenberg et al., 2006;Mannisto and Kaakkola, 1999]. The list of potential functional sites includes the following SNPs: rs2097603 in the promoter region of brain-expressed, membrane-bound (MB-) COMT form [Palmatier et al., 2004;Zhu et al., 2004;Chen et al., 2004], rs737865 upstream in the intron 1, and rs165599 in the 3′ untranslated region [Shifman et al., 2002;Bray et al., 2003]. Furthermore, in studying the association between COMT genotypes and variability in human pain perception, it has been found that the val158met polymorphism alone was not significantly associated with a derived measure of global pain sensitivity [Diatchenko et al., 2005]. Instead, three common haplotypes of the COMT gene, consisting of two synonymous (rs4633 and rs4818) and one nonsynonymous val158met SNPs are coding for different levels of enzymatic activity. Corresponding differences in pain sensitivity are associated with regulation of the translation efficiency through haplotype-dependent secondary mRNA structure [Nackley et al., 2006]. Thus, COMT contains at least five functional polymorphisms that potentially impact the index of pain sensitivity. Interactions of functional alleles at COMT imply that the genetic effects may not be easily inferred from the information on one SNP at a time, and that the SNP-specific effects may in fact be misleading. In our application, we tested four SNPs which have been previously independently associated with COMT-dependent phenotypes: rs2097603, rs737865, rs4680 and rs165599. An additional SNP rs4818 was chosen as a major contributor to functional COMT haplotypes that together with SNP rs4680 defines three functional haplotypes and influences pain sensitivity [Diatchenko et al., 2005]. All SNPs were found to conform to Hardy-Weinberg expectations. As a functional measure of COMT activity we used sensitivity to noxious stimuli. A summary measure of thermal threshold has been chosen for this study as it has been shown to be the most sensitive measure with respect to association with COMT

genotypes, and has a small associated measurement error [Diatchenko et al., 2005,2006, 2007].

Two SNPs, rs737865 and rs4818 with the corresponding LD correlation of $r = 0.5$, show a significant mean effect as well as an indication that variability might be different between individuals carrying different alleles. There is a good correspondence between the asymptotic and the resampled p-values for the mean and the variance tests (based on 50,000 trait value permutations). There is less correspondence for the combined multilocus tests, $T^{\mu, V}$, which could be attributed to insufficient closeness to normality, small sample size, and accumulation of errors from T^μ and T^V that constitute the combined statistic. The major alleles, A of rs737865, and G of rs4818 are associated with the same direction of the effect, as well as with the variance increase. Haplotypes and diplotypes that carry either AG or GC combinations of alleles show significance associated with the variance parameter. The genotypic/diplotypic tests were included in those cases where there was an indication of a contrast in variance at the allelic/ haplotypic test, to rule out a possible variance increase due to dominance effects. Consistent with previous publications [Diatchenko et al., 2005, 2006, 2007], homozygotes for AG haplotype display the highest pain thresholds and the lowest sensitivity to pain, while homozygotes for GC haplotype display the lowest pain thresholds and the highest sensitivity to pain. Among the four haplotypes with the frequency of at least 0.05 that include all five markers and carry the opposite effect alleles at rs737865 and rs4818, one shows a significant variance effect (haplotype AAGGA). Results found to be significant at the 5% level by the T^μ, were also found to be significant by the test T^{μ^*} (data not shown). However, there was less correspondence between the asymptotic and the resampled values for the T^{μ^*}. There is an issue of multiple testing that should be accounted for in the interpretation of these results. Nevertheless, prior findings and the significance of the overall linear model that includes additive effects of all five SNPs (p=0.013) suggest the existence of an association in the region. The haplotype AGCGA with the significant mean effect corresponds to a high pain sensitivity ("HPS") haplotype of COMT [Diatchenko et al., 2005] that codes for a low translation efficiency of COMT. This leads to a 25-fold lower enzymatic activity when compared with the low pain sensitivity haplotype, LPS [Nackley et al., 2006]. The LPS corresponds to the haplotype AAGGA of the present study. A significant variance contrast at rs737865 and an *increased* variance associated with diplotype classes that include the haplotype GC point to a possibility of additional unobserved functional variants that have not been typed in this study. In those cases where a haplotype (or a diplotype) of interest shows an associated *decreased* variance relative to the group that includes several other haplotypes, a possibility should be examined that the variance in the combined group is increased due to a mixture of effects associated with heterogeneous haplotypes in that group.

Parameter estimates, type-I error, and overall tests

When testing one haplotype at a time, a given haplotype h_i is contrasted against the rest of the haplotypes, collapsed into a combined group, $h_{\bar{i}}$. Population means and variances for additive, recessive, dominant, and diplotypic models of analysis can be written in terms of the population frequencies and effects of haplotypes. These theoretical values are given in Appendix 2, and can be compared to the output of our EM algorithm for unphased data. The hypothesis tests assume a normal distribution, however the test T^μ for comparison of haplotypic means is robust to deviations from normality. The parameter estimates appear to be precise for non-normally distributed traits. In this section, we assumed haplotype contributions to the phenotype to follow either a normal or a Gamma distribution. For a haplotype h_i in the Gamma model, its contribution is represented by a random variable $X_i \sim \text{Gamma}(\alpha_i, \beta_i)$, with parameters given in Table 1. Two haplotype contributions combine in an individual as $X_i + X_j$. A trait histogram for a sample under this model would show a decidedly skewed distribution. The results (Table

8) using samples of 300 individuals show good correspondence with theoretical values, calculated according to Appendix 2.

Frequencies of the first five COMT haplotypes, as given in Table 1, cover the range from 0.064 to 0.3. These haplotypes were used to evaluate the type-I error rates as a function of frequency. Results of the simulations, for different models of analysis, are shown in Table 9. Tests for the haplotypic and the dominant models of analysis maintain the nominal 5% error rate even for the rare haplotype (the last row in the table), and the error rate for the recessive model is the worst, for the tests that incorporate the variance. The $T^{\mu}{}^{*}$ was evaluated but not included into the table for the dominant and the recessive models. This test maintained type-I error rate for the dominant model. For the recessive model, it was approximately as liberal as that shown for the diplotypic model. These results may be explained based on the expected number of categories used in the comparisons. With $N = 500$, and the haplotype frequency $p = 0.064$, the expected number of haplotypes is $2Np = 64$. However, the diplotype and the recessive models are both estimating the frequency of the homozygote, with the expected number (Np^2) that is equal to only about two. On the other hand, $p = 0.3$ gives the expected number of 45 individuals, which appears to be sufficient. Based on these results, a permutational test is recommended to verify small asymptotic p-values. Performance of the overall tests is summarized in Table 10. The additive (haplotypic) model of analysis was used. We included results for the Haplotype Trend Regression (HTR) method [Zaykin et al., 2002] for comparison. The first, "HTR", column gives power for the overall HTR test as described in Zaykin et al. [2002]. The second column gives power for the Simes-based HTR, and these values are very similar to those for the Simes-based T^{μ} (fifth column). Before the application of the Simes test, both of these approaches compare the effect of each haplotype in turn with the weighted average for the rest of the haplotypes. We note that HTR tests for individual haplotypes in these simulations gave power that was nearly identical to the power of the haplotypic test T^{μ} (data not shown). For non-normal data, such values are given in Table 11, again showing similarity between the two tests. It is also expected that a multi-parameter haplotypic LRT (T^{μ}) would yield power values that are similar to those given in the first column of Table 10, as based on the form of the test (12) and on results from Appendix 1 of Zaykin et al. [2002]. Columns 2–7 of Table 10 give values for the Simes-based tests, where p-values that correspond to the first five most frequent haplotypes, listed in Table 1, are combined into the overall p-value. One exception is the last row (Setting 8). For this row, we used the FUSION data frequencies of 5-SNP haplotypes, as given in Table 1 of Lin [2004], and originally described in Valle et al. [1998]. Epstein and Satten [2003] estimated susceptibility effects for these haplotypes and found two of the haplotypes to be significantly associated (the first with an increased, the second with a decreased susceptibility). Therefore, for these two haplotypes we allocated effects that are in opposite directions from the baseline value.

The results can be summarized as follows. Predictably, the Simes-HTR gives power values that are close to the values for the overall (Simes) T^{μ}. The Simes used as an overall test is more powerful than a multi-parameter test when there are one or two haplotypes with effects that are distinct from a baseline value. When all haplotypes have distinct mean effects, the multi-parameter overall test is more powerful than the Simes, however power values for the two methods are more similar when frequencies of haplotypes included into analysis are similar to each other. The type-I error for the Simes test is well maintained. In the presence of multiple haplotypic effects with the common variance (Settings 6,7,8), some increase in the type-I error for the T^{V} may be expected because of the mixture of effects associated with the combined group ($h_{\bar{i}}$, Appendix 2). Since the combined group is expected to show an increased rather than decreased variance, a possible modification of the test is to consider a one-tailed hypothesis $V_i > V_{\bar{i}}$.

To investigate the effect of normality violation on the performance of the tests, we performed additional simulations under the Gamma model, using tests for a single haplotype (h_1 with the frequency 0.3034). The power simulations (Table 11) indicate that only the tests T^μ and HTR retain a proper type-I error. The other tests are sensitive to non-normality, however the Box-Cox transformation remedies this problem. When the values $\alpha_i\beta_i$ are the same for all haplotypes, the tests incorporating the haplotypic variance show a substantial gain in power, as expected, based on the fact that $E(X_i)=\alpha_i\beta_i$; $V(X_i)=\alpha_i\beta_i^2$. The relative performance of T^μ and T^{μ^*} is switched in settings 3 and 4: As we have noted, T^{μ^*} tends to provide an increased power when a tested haplotype (contrasted against the combined group) is associated with a smaller variance, and when its frequency is smaller than the frequency of the combined group.

Caution is needed regarding the interpretation of results obtained for the transformed data. Even under normality, the mean and the variance-specific tests are independent only under the complete null. There might be some inflation of the type-I error for these tests when the second parameter is heterogeneous between the haplotypes, and problem is likely to be exaggerated by both non-normality as well as by the transformation. Summarizing results across different simulations, we conclude that when variances are unequal, size of the test T^{μ^*} is closer to the nominal level than that of the T^μ. However, T^{μ^*} is more sensitive to small expected counts, and we recommend that its asymptotic p-values need to be verified by a follow-up with a permutational test. Size of the variance-specific test in the presence of heterogeneity between the means is close to the nominal level (e.g. $p_{B_1}=0$ and $p_{B_1}=1$ rows of Table 4). Interpretation of results from testing the combined hypothesis ($H_0^{\mu,V}$) is most straightforward. We note that the HTR method [Zaykin et al., 2002] and a similar diplotype-based method (DTR) [Luo et al., 2006] can be recommended for testing hypotheses concerning specifically the haplotypic or diplotypic means. These methods are robust against deviations from normality, and they retain proper size regardless of whether the haplotypic variances are unequal. In these simulations, the HTR showed power values that are almost identical to the haplotypic T^μ (Table 11).

Discussion

The association models that we considered here assume that some relevant variation is unknown to investigators. Although this incomplete knowledge impairs power to detect haplotypic effects, it also induces differences in trait variances between the scored haplotypes. We exploit this phenomenon in the proposed approach. Inclusion of the variance parameters into haplotype association methods for unphased data is useful for two reasons. Firstly, this adds power to detect associations under certain models, especially under models where known and unknown variants interact with each other. Secondly, a variance increase at a haplotype serves as an indication that the variant under study is either correlated, or epistatically interacts with additional unobserved polymorphisms. We conclude that tests that account for the haplotype- or diplotype-specific variance are useful for discovering complex associations with quantitative traits. Power can be gained while examining only a subset of SNPs that are either directly involved in joint multilocus genetic effects, or linked with functional variants via LD. Our approach is not a replacement for the conventional way of comparing haplotypic means. The number of tests is an issue, therefore at a hypothesis-generating stage, for example in a whole genome association analysis, a routine utilization of the conventional approach is more suitable. Testing additional hypotheses that involve haplotypic and diplotypic variances is more relevant in a smaller, or a follow-up study, concerned with a number of preselected loci.

The unknown factor influencing the trait can be non-genetic. Both, the usual type of an associations test for comparison of the haplotypic effects, as well as a test that incorporates haplotype-specific variances are sensitive to confounding. However, in the absence of

confounding, differences between either the haplotypic means, or the variances in this scenario would indicate a partial functional involvement of the scored polymorphisms in the association with the trait, as well as an interaction with the unknown factor.

The approach outlined here is applicable to random samples from homogeneous populations, and thus, it is subject to the usual concerns about possibilities of confounding and stratification. Recently, Epstein et al. [2007] described a simple and effective method to correct for population stratification that can be used in conjunction with our approach. Further extensions are possible that will incorporate family-based samples.

A priori, untyped polymorphisms that reside within the same gene might be considered to be the first candidates that may account for a variance contrast between haplotypes. This is because models without the LD are "less general", in that they require a particular form of epistasis (or, equivalently, a particular form of interaction with an environmental factor) in order to induce a haplotypic variance contrast [Zaykin and Shibata, 2008]. A gain in power for the proposed method is higher under models that involve some degree of interaction with or without LD. Considering that compensatory changes that follow functional mutations are likely to be a ubiquitous force in the genome evolution [Kern and Kondrashov, 2004; Kirby et al., 1995; Kondrashov et al., 2002], multiple interacting functional SNPs within a gene locus could be relatively common in the haplotypic organization of human genome. How easily these interactions and the corresponding genetic variants might be identified remains an open question, given statistical difficulties related to detection and characterization of complex multilocus effects.

Software availability

Software implementing statistical approaches described here is available from the authors upon request. Simulation scripts used in this study are available upon request from DVZ.

Acknowledgments

References

Allen AS, Satten GA. Robust estimation and testing of haplotype effects in case-control studies. Genet Epidemiol 2008;32:29–40. [PubMed: 17948229]

Balding DJ. Likelihood-based inference for genetic correlation coefficients. Theor Popul Biol 2003;63:221–230. [PubMed: 12689793]

Bartlett M. Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences 1937;160:268–282.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc B 1995;57:289–300.

Box G, Cox D. An Analysis of Transformations. J Royal Stat Soc B 1964;26:211–252.

Bray NJ, Buckland PR, Williams NM, Williams HJ, Norton N, Owen MJ, O'Donovan MC. A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. Am J Hum Genet 2003;73:152–161. [PubMed: 12802784]

Chen J, Lipska BK, Halim N, Ma QD, Matsumoto M, Melhem S, Kolachana BS, Hyde TM, Herman MM, Apud J, Egan MF, Kleinman JE, Weinberger DR. Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. Am J Hum Genet 2004;75:807–821. [PubMed: 15457404]

Diatchenko L, Nackley AG, Slade GD, Belfer I, Max MB, Goldman D, Maixner W. Responses to Drs. Kim and Dionne regarding comments on Diatchenko, et al. Catechol-O-methyltransferase gene polymorphisms are associated with multiple pain-evoking stimuli. Pain 2006;125:216–24. Pain 2007;129:366–370. [PubMed: 17851590]

Diatchenko L, Nackley AG, Slade GD, Bhalang K, Belfer I, Max MB, Goldman D, Maixner W. Catechol-O-methyltransferase gene polymorphisms are associated with multiple pain-evoking stimuli. Pain 2006;125:216–224. [PubMed: 16837133]

Diatchenko L, Slade GD, Nackley AG, Bhalang K, Sigurdsson A, Belfer I, Goldman D, Xu K, Shabalina SA, Shagin D, Max MB, Makarov SS, Maixner W. Genetic basis for individual variations in pain perception and the development of a chronic pain condition. Hum Mol Genet 2005;14:135–143. [PubMed: 15537663]

Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE, Goldman D, Weinberger DR. Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. Proc Natl Acad Sci USA 2001;98:6917–6922. [PubMed: 11381111]

Elston R. On Fisher's method of combining p-values. Biometrical journal 1991;33:339–345.

Enoch MA, Xu K, Ferro E, Harris CR, Goldman D. Genetic origins of anxiety in women: a role for a functional catechol-O-methyltransferase polymorphism. Psychiatr Genet 2003;13:33–41. [PubMed: 12605099]

Epstein M, Satten G. Inference on haplotype effects in case-control studies using unphased genotype data. Am J Hum Genet 2003;73:1316–1329. [PubMed: 14631556]

Epstein MP, Allen AS, Satten GA. A simple and improved correction for population stratification in case-control studies. Am J Hum Genet 2007;80:921–930. [PubMed: 17436246]

Karayiorgou M, Sobin C, Blundell ML, Galke BL, Malinova L, Goldberg P, Ott J, Gogos JA. Family-based association studies support a sexually dimorphic effect of COMT and MAOA on genetic susceptibility to obsessive-compulsive disorder. Biol Psychiatry 1999;45:1178–1189. [PubMed: 10331110]

Kern AD, Kondrashov FA. Disease-related versus polymorphic mutations in human mitochondrial tRNAs. Nat Genet 2004;36:1207–1212. [PubMed: 15502829]

Kirby DA, Muse SV, Stephan W. Maintenance of pre-mRNA secondary structure by epistatic selection. Proceedings of the National Academy of Sciences 1995;92:9047–9051.

Kondrashov AS, Sunyaev S, Kondrashov FA. Dobzhansky-Muller incompatibilities in protein evolution. Proc Natl Acad Sci USA 2002;99:14878–14883. [PubMed: 12403824]

Kraft P, Cox D, Paynter R, Hunter D, De Vivo I. Accounting for haplotype uncertainty in matched association studies: A comparison of simple and flexible techniques. Genetic Epidemiology 2005;28:261–272. [PubMed: 15637718]

Kraft P, Stram D. Re: The use of inferred haplotypes in downstream analysis. Am J Hum Genet 2007;81:863–865. [PubMed: 17847013]

Lin D. Haplotype-based association analysis in cohort studies of unrelated individuals. Genetic Epidemiology 2004;26:255–264. [PubMed: 15095385]

Lin D, Huang B. Reply to Peter Kraft and Daniel O. Stram. Am J Hum Genet 2007;81:865–866.

Lin D, Zeng D. Likelihood-based inference on haplotype effects in genetic association studies. J Am Stat Assoc 2006;101:89–104.

Lin D, Zeng D, Millikan R. Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. Genet Epidemiol 2005;29:299–312. [PubMed: 16240443]

Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. Am J Hum Genet 2007;80:531–538. [PubMed: 17273975]

Lotta T, Vidgren J, Tilgmann C, Ulmanen I, Meln K, Julkunen I, Taskinen J. Kinetics of human soluble and membrane-bound catechol O-methyltransferase: a revised mechanism and description of the thermolabile variant of the enzyme. Biochemistry 1995;34:4202–4210. [PubMed: 7703232]

Luo X, Kranzler H, Zuo L, Wang S, Schork N, Gelernter J. Diplotype trend regression analysis of the *ADH* gene cluster and the *ALDH2* gene: Multiple significant associations with alcohol dependence. Am J Hum Genet 2006;78:973–987. [PubMed: 16685648]

Mannisto PT, Kaakkola S. Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. Pharmacol Rev 1999;51:593–628. [PubMed: 10581325]

Meyer-Lindenberg A, Nichols T, Callicott JH, Ding J, Kolachana B, Buckholtz J, Mattay VS, Egan M, Weinberger DR. Impact of complex genetic variation in COMT on human brain function. Mol Psychiatry 2006;11:867–877. [PubMed: 16786032]

Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS, Maixner W, Di-atchenko L. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science 2006;314:1930–1933. [PubMed: 17185601]

Nackley AG, Tan KS, Fecho K, Flood P, Diatchenko L, Maixner W. Catechol-O-methyltransferase inhibition increases pain sensitivity through activation of both beta2- and beta3-adrenergic receptors. Pain 2007;128:199–208. [PubMed: 17084978]

Nielsen DM, Weir BS. A classical setting for associations between markers and loci affecting quantitative traits. Genet Res 1999;74:271–277. [PubMed: 10689804]

Oroszi G, Goldman D. Alcoholism: genes and mechanisms. Pharmacogenomics 2004;5:1037–1048. [PubMed: 15584875]

Palmatier MA, Pakstis AJ, Speed W, Paschou P, Goldman D, Odunsi A, Okonofua F, Kajuna S, Karoma N, Kungulilo S, Grigorenko E, Zhukova OV, Bonne-Tamir B, Lu RB, Parnas J, Kidd JR, DeMille MMC, Kidd KK. COMT haplotypes suggest P2 promoter region relevance for schizophrenia. Mol Psychiatry 2004;9:859–870. [PubMed: 15098000]

Schaid D, Rowland C, Tines D, Jacobson R, Poland G. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 2002;70:425–434. [PubMed: 11791212]

Shi MM, Bleavins MR, de la Iglesia FA. Technologies for detecting genetic polymorphisms in pharmacogenomics. Mol Diagn 1999;4:343–351. [PubMed: 10671645]

Shibata K, Ito T, Kitamura Y, Iwasaki N, Tanaka H, Kamatani N. Simultaneous estimation of haplotype frequencies and quantitative trait parameters: applications to the test of association between phenotype and diplotype configuration. Genetics 2004;168:525–539. [PubMed: 15454562]

Shifman S, Bronstein M, Sternfeld M, Pisant-Shalom A, Lev-Lehman E, Weizman A, Reznik I, Spivak B, Grisaru N, Karp L, Schiffer R, Kotler M, Strous RD, Swartz-Vanetik M, Knobler HY, Shinar E, Beckmann JS, Yakir B, Risch N, Zak NB, Darvasi A. A highly significant association between a COMT haplotype and schizophrenia. Am J Hum Genet 2002;71:1296–1302. [PubMed: 12402217]

Simes R. An improved Bonferroni procedure for multiple tests of significance. Biometrika 1986;73:751–754.

Stram D, Leigh Pearce C, Bretsky P, Freedman M, Hirschhorn J, Altshuler D, Kolonel L, Henderson B, Thomas D. Modeling and E-M estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals. Hum Hered 2003;55:179–190. [PubMed: 14566096]

Tregouet D, Barbaux S, Escolano S, Tahri N, Golmard J, Tiret L, Cambien F. Specific haplotypes of the P-selectin gene are associated with myocardial infarction. Human Molecular Genetics 2002;11:2015–2023. [PubMed: 12165563]

Tzeng J, Wang C, Kao J, Hsiao C. Regression-based association analysis with clustered haplotypes through use of genotypes. Am J Hum Genet 2006;78:231–242. [PubMed: 16365833]

Valle T, Tuomilehto J, Bergman RN, Ghosh S, Hauser ER, Eriksson J, Nylund SJ, Kohtamaki K, Toivanen L, Vidgren G, Tuomilehto-Wolf E, Ehnholm C, Blaschak J, Langefeld CD, Watanabe RM, Magnuson V, Ally DS, Hagopian WA, Ross E, Buchanan TA, Collins F, Boehnke M. Mapping genes for NIDDM. Design of the Finland-United States investigation of NIDDM genetics (FUSION) study. Diabetes Care 1998;21:949–958. [PubMed: 9614613]

Weir, BS. Genetic data analysis II. Sinauer Associates Sunderland; Mass: 1996.

Wright S. The genetical structure of populations. Ann Eugen 1951;15:323–354.

Xie R, Stram D. Asymptotic equivalence between two score tests for haplotype-specific risk in general linear models. Genet Epidemiol 2005;29:166–70. [PubMed: 16025443]

Zaykin DV, Shibata K. Genetic flip-flop without an accompanying change in linkage disequilibrium. Am J Hum Genet 2008;82:794–796. [PubMed: 18319078]

Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 2002;53:79–91. [PubMed: 12037407]

Zaykin DV, Zhivotovsky LA, Czika W, Shao S, Wolfinger RD. Combining p-values in large-scale genomics experiments. Pharm Stat 2007;6:217–226. [PubMed: 17879330]

Zhao L, Li S, Khalid N. A method for the assessment of disease associations with single-nucleotide polymorphism haplotypes and environmental variables in case-control studies. Am J Hum Genet 2003;72:1231–1250. [PubMed: 12704570]

Zhu G, Lipsky RH, Xu K, Ali S, Hyde T, Kleinman J, Akhtar LA, Mash DC, Goldman D. Differential expression of human COMT alleles in brain and lymphoblasts detected by RT-coupled 5′ nuclease assay. Psychopharmacology 2004;177:178–184. [PubMed: 15290009]

Zubieta JK, Heitzeg MM, Smith YR, Bueller JA, Xu K, Xu Y, Koeppe RA, Stohler CS, Goldman D. COMT val158met genotype affects mu-opioid neurotransmitter responses to a pain stressor. Science 2003;299:1240–1243. [PubMed: 12595695]

Appendix 1. Effects at the scored locus under the heterogeneity model

Following Nielsen and Weir [1999], the effect of a scored allele, A_1, can be written as

$$\mu_{A_1} = \sum_i E(Y|B_iA_1)\Pr(B_i|A_1)$$
$$= \sum_i E(Y|B_i)\Pr(B_i|A_1)$$
$$= \sum_i \mu_{B_i}\Pr(B_i|A_1) \tag{15}$$

The second equality holds because A is now only a marker for nearby functional variants. The probability $\Pr(B_i \mid A_1)$ is a function of LD, i.e. $\Pr(B \mid A_1) = p_B + D/p_{A_1}$, where D is the LD coefficient, $p_{A_1B_i} - p_{A_1}p_{B_i}$. Consider the effect difference

$$\mu_{A_1} - \mu_{A_2} = \sum_i \mu_{B_i}[\Pr(B_i|A_1) - \Pr(B_i|A_2)]$$
$$= \sum_i \mu_{B_i}\left[\frac{p_{B_iA_1}}{p_{A_1}} - \frac{p_{B_iA_2}}{1-p_{A_1}}\right]$$

The difference is zero under linkage equilibrium. Otherwise, the sign of the effect difference depends on the haplotype frequencies, $p_{B_iA_1}, p_{B_iA_2}$. The association effects and the LD coefficients can effectively cancel each other at the observed "proxy" locus, yet the variance contrast may remain substantial.

In the simulation experiments, the joint frequencies $p_{B_iA_j}$ were modeled assuming a Dirichlet distribution. This distribution of population haplotype frequencies is justified under certain population genetic models, such as models of drift-mutation equilibrium, as well as under some non-equilibrium models [Wright, 1951; Balding, 2003]. Ten randomly chosen Dirichlet parameters were set to 50, while the remaining 10 parameters were set to 1/20. By rejection sampling, the frequency p_{A_1} was ensured to be in the interval [0.27, 0.33], to approximate the frequency of the common COMT haplotype. This model allocates an expected frequency of around 0.09 to ten of the haplotypes, while generating a tail of rare haplotypes. The average LD correlation across simulations was $r_{B_iA_1} = 0.2$, and the 95% quantile for the correlation was 0.55. Common variance of the trait was assumed for all functional variants, $\sigma^2 = 1$. Assuming the underlying normal distribution and additivity of allelic effects, the effect means were taken to be 1/2, 1,…,5. To conclude a single simulation experiment, a random sample of 500 diploid individuals was taken from the population constructed as just described. The

advantage of generating new population haplotype frequencies for each simulation run is in that it allows us to examine the expected behavior of statistical tests over a range of different haplotype frequency configurations.

Appendix 2. Population parameter values

Here we consider theoretical parameter values for the mean and the standard deviation given additive, diplotype, dominant and recessive models of analysis. We assume the population HWE. For the purpose of evaluating the estimation techniques, we assume that the two haplotypic contributions, represented by random variables $X_i. X_j$, combine in a diplotype as $X_i + X_j$, discarding dominance deviations. We denote the mean and the variance of X_i by θ_i and σ_i^2, correspondingly. Thus, such model assumes additivity of haplotypic effects θ_i, θ_j in a diplotype, $\theta_{ij} = \theta_i + \theta_j$. This allows to simplify the formulas, however more general expressions that keep notation in terms of the diplotype effects and their frequencies are obtained similarly. The population mean value and the trait variance in this model are

$$\bar{\mu} = \sum_i^k \sum_j^k [p_{ij}(\theta_i + \theta_j)]$$
$$= \sum_i^k \sum_j^k [p_i p_j(\theta_i + \theta_j)]$$
$$= 2 \sum_i^k p_i \theta_i$$
$$\bar{V} = \sum_i^k \sum_j^k p_i p_j [\sigma_i^2 + \sigma_j^2 + (\theta_i + \theta_j - \bar{\mu})^2]$$

where p_i is the frequency of the haplotype h_i, $\{i = 1, \ldots, k\}$.

Additive (haplotypic) model

Additive effect of the haplotype h_i is defined as the effect of a diplotype that contains the h_i, while the second haplotype is randomly drawn from the same population. The variance is defined similarly. From this definition, the additive mean and the variance for the haplotype h_i are

$$\mu_i = \theta_i + \bar{\mu}/2$$
$$V_i = \sigma_i^2 + \sum_j^k p_j \sigma_j^2 + (\theta_j - \bar{\mu}/2)^2$$

For the rest of the haplotypes, collapsed into a composite group, $h_{\bar{i}}$, the values are

$$\mu_{\bar{i}} = \mu_a + \bar{\mu}/2$$
$$V_{\bar{i}} = \sum_j^k q_j \sigma_j^2 + (\theta_j - \mu_a)^2 + \sum_i^k p_i \sigma_i^2 + (\theta_i - \bar{\mu}/2)^2$$

where

$$q_i = \frac{p_i}{\sum_{j \neq i}^{k} p_j}$$

$$\mu_a = \sum_{j \neq i}^{k} q_j \theta_j$$

Diplotype model

The diplotype model distinguishes between the three diplotype classes, h_i/h_i, $h_i/h_{\bar{i}}$, $h_{\bar{i}}/h_{\bar{i}}$. The means and the variances are

$$\mu_{ii} = 2\theta_i$$
$$\mu_{i\bar{i}} = \theta_i + \mu_a$$
$$\mu_{\bar{i}\bar{i}} = 2\mu_a$$
$$V_{ii} = 2\sigma_i^2$$
$$V_{i\bar{i}} = \sigma_i^2 + \sum_{j \neq i}^{k} q_j \sigma_j^2 + (\theta_j - \mu_a)^2$$
$$V_{\bar{i}\bar{i}} = \sum_{s \neq i}^{k} \sum_{t \neq i}^{k} q_s q_t [\sigma_s^2 + \sigma_t^2 + (\theta_s + \theta_t - 2\mu_a)^2]$$

Recessive model

The recessive model distinguishes between the two diplotype classes, h_i/h_i, and ($h_i/h_{\bar{i}}$ or $h_{\bar{i}}/h_{\bar{i}}$). The means and the variances are

$$\mu_{ii} = 2\theta_i$$
$$\mu_{i\bar{i}/\bar{i}\bar{i}} = \frac{\sum_s^k \sum_t^k p_s p_t (\theta_s + \theta_t)\{1 - I(s=t=i)\}}{1 - p_i^2}$$
$$V_{ii} = 2\sigma_i^2$$
$$V_{i\bar{i}/\bar{i}\bar{i}} = \frac{\sum_s^k \sum_t^k p_s p_t [\sigma_s^2 + \sigma_t^2 + (\theta_s + \theta_t - \mu_{i\bar{i}/\bar{i}\bar{i}})^2]\{1 - I(s=t=i)\}}{1 - p_i^2}$$

where $I(\cdot)$ is the indicator function.

Dominant model

The dominant model distinguishes between the two diplotype classes, (h_i/h_i or $h_i/h_{\bar{i}}$), and $h_{\bar{i}}/h_{\bar{i}}$. The means and the variances are

$$\mu_{ii/i\bar{i}} = \frac{\left[2\sum_j^k p_i p_j (\theta_i + \theta_j)\right] - 2\theta_i p_i^2}{1 - \sum_{s \neq i}^{k} \sum_{t \neq i}^{k} p_s p_t}$$
$$V_{ii/i\bar{i}} = \frac{\left[2\sum_j^k p_i p_j \{\sigma_i^2 + \sigma_j^2 + (\theta_i + \theta_j - \mu_{ii/i\bar{i}})^2\}\right] - p_i^2 \{2\sigma_i^2 + (2\theta_i - \mu_{ii/i\bar{i}})^2\}}{1 - \sum_{s \neq i}^{k} \sum_{t \neq i}^{k} p_s p_t}$$
$$\mu_{\bar{i}\bar{i}} = 2\mu_a$$
$$V_{\bar{i}\bar{i}} = \sum_i^k \sum_j^k q_i q_j [\sigma_i^2 + \sigma_j^2 + (\theta_i + \theta_j - 2\mu_a)^2]$$
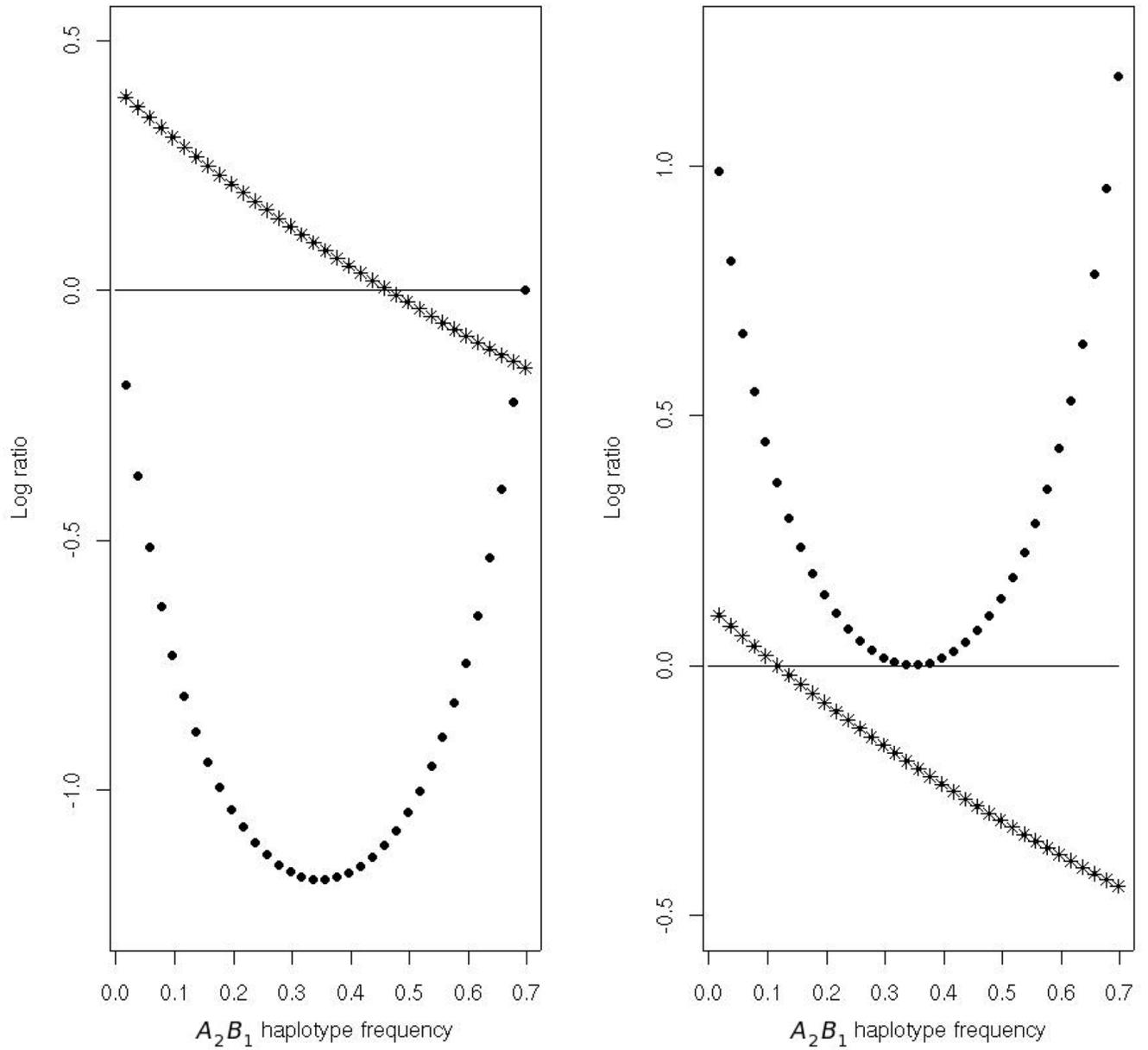
Figure 1.
Values for $\ln(\mu_{A_1}/\mu_{A_2})$ (*) and $\ln(V_{A_1}/V_{A_2})$ ($\bullet$) for the non-interactive model; $\sigma^2 = 1$. Left graph, (a): $p_{A_1B_1} = 0.3034$; $p_{A_1B_2} = 0$. Right graph, (b): $p_{A_1B_1} = p_{A_1B_2} = 0.3034/2$.
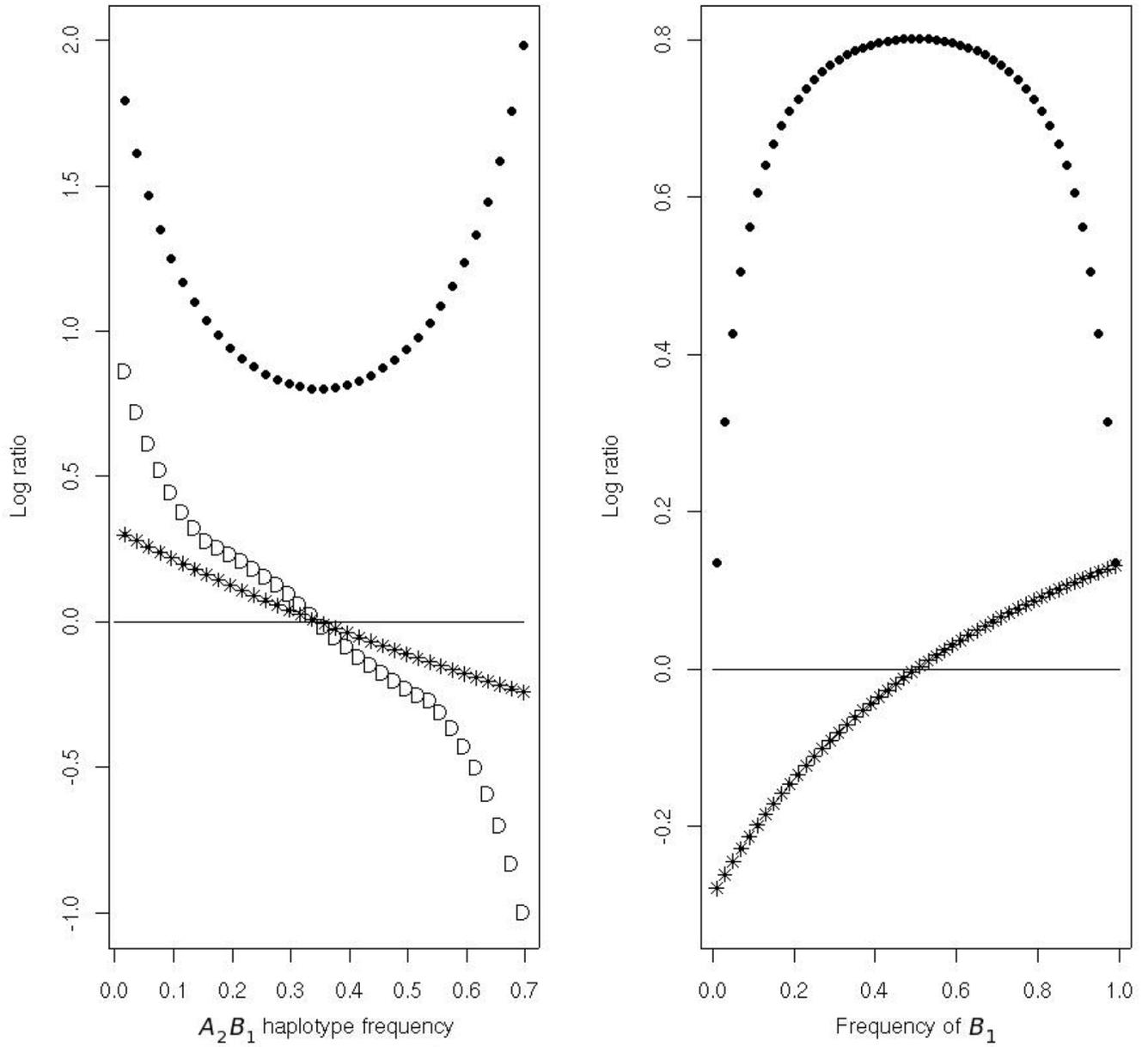
Figure 2.
Values for $\ln(\mu_{A_1}/\mu_{A_2})$ (*) and $\ln(V_{A_1}/V_{A_2})$ (●) for the interactive model; $\sigma^2 = 1$. Left graph,
(a): $p_{A_1B_1} = p_{A_1B_2} = 0.3034/2$; "D" denotes the standardized LD (D'). Right graph, (b): Linkage
equilibrium.
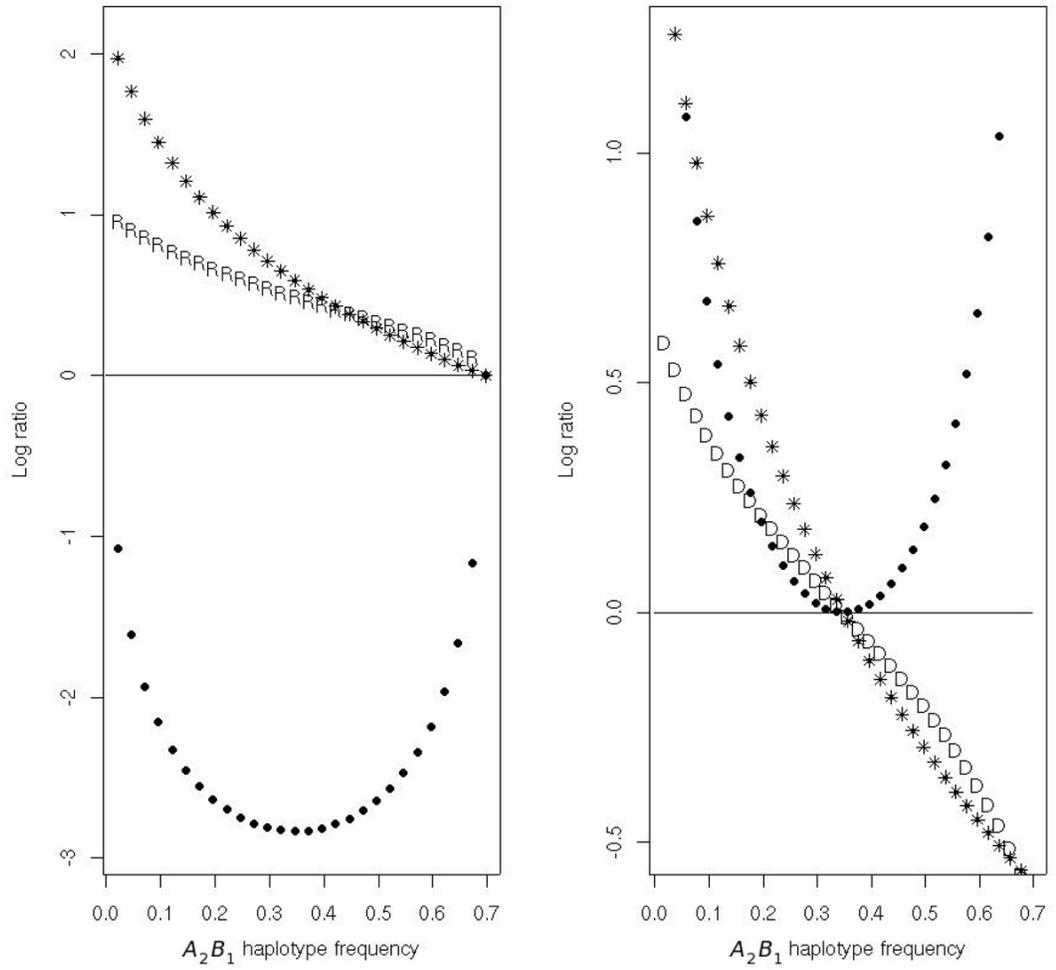
Figure 3.
Values for $\ln(\mu_{A_1}/\mu_{A_2})$ (*) and $\ln(V_{A_1}/V_{A_2})$ (●) for the LD proxy model; $\sigma^2 = 1$. Left graph, (a): $p_{A_1B_1} = 0.3034$; $p_{A_1B_2} = 0$; "R" denotes the LD correlation. Right graph, (b): $p_{A_1B_1} = p_{A_1B_2} = 0.3034/2$; "D" denotes the standardized LD (D').

Table 1

COMT Haplotype frequencies and parameters used in simulation experiments

Haplotype	Frequency	Shape[†]	Scale[†]
h_1 (AGCG)	0.3034	1	1.0
h_2 (AATC)	0.2909	2	1.2
h_3 (GATC)	0.2264	3	1.4
h_4 (GGCG)	0.0967	4	1.5
h_5 (AACC)	0.0637	5	1.6
h_6 (GACC)	0.0138	6	1.7
h_7(AGTC)	0.0029	7	1.8
h_8 (AACG)	0.0022	8	2.0

[†]
Shape (α) and scale (β) refer to parameters of the gamma distribution used for the haplotypic effect and variance estimation simulations.

Haplotypes consist of the following COMT SNPs: rs2075507, rs6269, rs4633, and rs4818 with the pairwise LD correlations [Weir, 1996] equal to 0.271, −0.335, −0.270, −0.844, and −0.989.

Table 2

LRT power values under the non-interactive model

$P_{A_2B_1}$; σ^2	Allelic (SNP)			Haplotypic					Diplotypic			
	$T^{\mu,V}$	T^μ	T^V	$T^{\mu,V}$	$T^{\mu*,V}$	T^μ	$T^{\mu*}$	T^V	$T^{\mu,V}$	T^μ	$T^{\mu*}$	T^V
0; 50	0.728	0.818	0.052	0.718	0.727	0.810	0.816	0.052	0.612	0.725	0.743	0.054
0.2; 10	0.885	0.912	0.192	0.875	0.885	0.901	0.913	0.181	0.792	0.833	0.867	0.153
0.25; 5	0.948	0.920	0.497	0.945	0.953	0.919	0.935	0.489	0.902	0.858	0.904	0.433
0.35; 2	0.971	0.610	0.945	0.968	0.969	0.601	0.684	0.942	0.953	0.444	0.669	0.937
0.4644; 2	0.858	0.032	0.924	0.844	0.848	0.034	0.057	0.918	0.786	0.024	0.063	0.899
0.55; 2	0.905	0.464	0.821	0.889	0.899	0.449	0.520	0.801	0.826	0.315	0.468	0.763
0.6; 2	0.972	0.898	0.629	0.970	0.971	0.890	0.915	0.627	0.937	0.809	0.882	0.574
0.65; 5	0.890	0.924	0.121	0.881	0.885	0.918	0.925	0.119	0.805	0.860	0.881	0.109
0.6966; 15	0.650	0.749	0.054	0.635	0.643	0.739	0.741	0.054	0.525	0.639	0.663	0.056

Model: $\mathbf{M} = \{\kappa_{A_1} + \kappa_{B_1}, \kappa_{A_1} + \kappa_{B_2}, \kappa_{A_2} + \kappa_{B_1}, \kappa_{A_2} + \kappa_{B_2}\}$; where $\kappa_{A_1} = 2$, $\kappa_{A_2} = 3$, $\kappa_{B_1} = 4$, $\kappa_{B_2} = 1$; $\mu_{A_1} = \mu_{A_2}$ at $p_{A_2B_1} = 0.4644$; $V_{A_1} = V_{A_2}$ at $p_{A_2B_1} = 0$ and 0.6966; the corresponding graph is Figure 1(a).

Table 3

LRT power values under the interactive model and LD

$P_{A_2B_1}$; σ^2	Allelic (SNP)			Haplotypic					Diplotypic			
	$T^{\mu,V}$	T^μ	T^V	$T^{\mu,V}$	$T^{\mu^*,V}$	T^μ	T^{μ^*}	T^V	$T^{\mu,V}$	T^μ	T^{μ^*}	T^V
0; 20	0.895	0.870	0.414	0.893	0.885	0.860	0.850	0.412	0.831	0.804	0.796	0.326
0.1; 10	0.888	0.779	0.611	0.884	0.881	0.768	0.754	0.602	0.825	0.705	0.695	0.501
0.2; 5	0.833	0.493	0.754	0.833	0.822	0.488	0.453	0.755	0.755	0.429	0.404	0.653
0.3483; 2	0.826	0.075	0.894	0.822	0.813	0.070	0.056	0.898	0.727	0.097	0.071	0.816
0.5; 5	0.843	0.508	0.765	0.844	0.831	0.494	0.472	0.763	0.771	0.444	0.420	0.661
0.6; 10	0.894	0.793	0.617	0.892	0.887	0.784	0.766	0.622	0.894	0.793	0.708	0.617
0.6966; 20	0.898	0.870	0.420	0.889	0.885	0.862	0.848	0.414	0.831	0.800	0.795	0.331

Model: $M = \{\varepsilon + \kappa_{A_1} + \kappa_{B_1}, \; \kappa_{A_1} + \kappa_{B_2}, \; \kappa_{A_2} + \kappa_{B_1}, \; \kappa_{A_2} + \kappa_{B_2}\}$; where $\varepsilon = 2$, $\kappa_{A_1} = 2$, $\kappa_{A_2} = 3$, $\kappa_{B_1} = 4$, $\kappa_{B_2} = 1$; $\mu_{A_1} = \mu_{A_2}$ at $p_{A_2B_1} = 0.3483$; the corresponding graph is Figure 2(a).

Table 4

LRT power values under the interactive model and LE

p_{B_1}; σ^2	Allelic (SNP)			Haplotypic					Diplotypic			
	$T^{\mu,V}$	T^μ	T^V	$T^{\mu,V}$	$T^{\mu*,V}$	T^μ	T^*	T^V	$T^{\mu,V}$	T^μ	$T^{\mu*}$	T^V
0; 10	0.822	0.889	0.051	0.820	0.819	0.890	0.888	0.054	0.727	0.825	0.833	0.053
0.1; 2	0.994	0.952	0.551	0.994	0.992	0.950	0.934	0.518	0.983	0.923	0.930	0.526
0.2; 2	0.958	0.665	0.750	0.949	0.943	0.658	0.602	0.730	0.910	0.592	0.581	0.676
0.3; 2	0.887	0.332	0.837	0.882	0.874	0.318	0.280	0.836	0.810	0.286	0.274	0.760
0.4; 2	0.839	0.124	0.879	0.833	0.825	0.127	0.105	0.878	0.754	0.143	0.113	0.806
0.5; 2	0.825	0.070	0.897	0.814	0.808	0.072	0.054	0.888	0.735	0.099	0.070	0.813
0.6; 2	0.838	0.135	0.882	0.829	0.827	0.125	0.107	0.876	0.740	0.133	0.114	0.795
0.7; 2	0.894	0.322	0.847	0.885	0.874	0.319	0.284	0.836	0.811	0.288	0.273	0.762
0.8; 2	0.956	0.663	0.747	0.952	0.944	0.655	0.604	0.738	0.910	0.586	0.586	0.679
0.9; 2	0.995	0.956	0.546	0.994	0.993	0.953	0.933	0.519	0.986	0.932	0.928	0.522
1; 10	0.825	0.894	0.056	0.815	0.817	0.886	0.885	0.057	0.723	0.819	0.830	0.057

Model: $M = \{\varepsilon + \kappa_{A_1} + \kappa_{B_1}; \kappa_{A_1} + \kappa_{B_2}; \kappa_{A_2} + \kappa_{B_1}; \kappa_{A_2} + \kappa_{B_2}\}$; where $\varepsilon = 2$, $\kappa_{A_1} = 2$, $\kappa_{A_2} = 3$, $\kappa_{B_1} = 4$, $\kappa_{B_2} = 1$; $\mu_{A_1} = \mu_{A_2}$ at $p_{A_2B_1} = 0.5$; the corresponding graph is Figure 2(b).

Table 5

LRT power values under the LD proxy model

$P_{A_2B_1}$; σ^2	Allelic (SNP)			Haplotypic					Diplotypic			
	$T^{\mu,\nu}$	T^μ	T^ν	$T^{\mu,\nu}$	$T^{\mu*,\nu}$	T^μ	$T^{\mu*}$	T^ν	$T^{\mu,\nu}$	T^μ	$T^{\mu*}$	T^ν
0.35; 100	0.937	0.962	0.166	0.925	0.934	0.950	0.959	0.163	0.869	0.907	0.932	0.136
0.45; 70	0.839	0.865	0.224	0.830	0.841	0.852	0.872	0.218	0.743	0.773	0.818	0.184
0.5; 20	0.984	0.970	0.727	0.980	0.983	0.967	0.976	0.727	0.970	0.929	0.968	0.691
0.55; 20	0.954	0.899	0.765	0.946	0.954	0.885	0.916	0.755	0.930	0.798	0.900	0.740
0.6; 10	0.897	0.730	0.786	0.878	0.887	0.712	0.764	0.776	0.864	0.572	0.734	0.776
0.65; 4	0.812	0.420	0.817	0.800	0.809	0.410	0.504	0.820	0.847	0.261	0.476	0.881

Model: $\mathbf{M} = \{9, 1, 9, 1\}$; the corresponding graph is Figure 3(a).

Table 6

LRT power values under the heterogeneity model (10 mutations, $B_1,...,B_{10}$).

Analysis model	Hypothesis					
	$T^{\mu,V}$	$T^{\mu *V}$	T^μ	$T^{\mu *}$	T^V	
Additive (SNP)	0.788	0.788	0.705	0.711	0.462	
Additive (SNP)[†]	0.799	0.802	0.705	0.705	0.493	
Additive (haplotypes)	0.771	0.773	0.685	0.696	0.422	
Genotypic (SNP)	0.745	0.750	0.667	0.682	0.418	
Diplotypic (haplotypes)	0.719	0.721	0.654	0.663	0.384	
Dominant (haplotypes)	0.718	0.718	0.655	0.655	0.337	
Recessive (haplotypes)	0.644	0.654	0.587	0.613	0.361	
Additive (haplotypes), H_0	0.050	0.050	0.051	0.051	0.049	
Diplotypic (haplotypes), H_0	0.048	0.057	0.049	0.056	0.049	
Dominant (haplotypes), H_0	0.047	0.047	0.050	0.049	0.045	
Recessive (haplotypes), H_0	0.054	0.059	0.050	0.058	0.055	

[†]
test based on a resampling distribution

Table 7

Association of COMT SNPs with the thermal threshold phenotype – effect estimates and LRT p-values.

SNP	μ_A	μ_a	V_A	V_a	$H_0^{\mu,V}$	H_0^{μ}	H_0^{V}
						p-values (asymptotic/resampled)	
m1	−0.051	0.078	0.911	1.110	0.214/0.197	0.233/0.221	0.198/0.194
m2	−0.167	0.084	0.777	1.083	0.010/0.019	0.026/0.025	0.042/0.043
m2 ^{G}	−0.001,−0.514	0.128	0.900,0.340	1.171	0.003/0.015	0.040/0.042	0.010/0.022
m3	0.124	−0.177	1.057	0.851	0.010/0.014	0.005/0.004	0.161/0.157
m4	−0.075	0.072	0.932	1.044	0.288/0.255	0.166/0.151	0.451/0.445
m5	0.012	−0.006	0.964	1.009	0.949/0.938	0.878/0.877	0.776/0.776
m2–m3 †	−0.196	0.073	0.740	1.069	0.0042/0.008	0.021/0.014	0.018/0.017
m2–m3 ^{†G}	−0.146,−0.412	0.168	0.855,0.184	1.132	0.003/0.018	0.045/0.047	0.007/0.027
m2–m3 ‡	0.144	−0.158	1.071	0.863	0.0048/0.014	0.0037/0.0038	0.135/0.153
m2–m3 ^{‡G}	0.111,0.173	−0.444	0.994,1.136	0.566	7.0e-04/0.005	0.0027/0.003	0.025/0.039
m1–...–m5 $^{\#}$	0.496	−0.024	1.123	0.976	0.077/0.146	0.026/0.034	0.687/0.725
m1–...–m5 ^{b}	−0.269	0.045	0.507	1.061	7.3e-05/9.0e-04	0.028/0.022	1.6e-04/7.4e-03
m1–...–m5 ^{b,G}	−0.240,−0.546	0.096	0.529,0.209	1.139	4.6e-04/0.005	0.072/0.070	5.7e-04/0.006

The tests are allele/haplotype-based unless indicated by a superscript G that stands for a genotype/diplotype model. In this case, the entries under μ_A, V_A refer to diplotypes carrying 1 and 2 copies of the haplotype, correspondingly. The entries under μ_a, V_a refer to diplotypes with 0 copies of the haplotype. SNP abbreviations: m1=rs2097603 (A>G); m2=rs737865 (A>G); m3=rs4818 (C>G); m4=rs4680 (G>A), m5=rs165599 (A>G).

†Haplotype AG.

‡Haplotype GC.

$^{\#}$Haplotype AGCGA.

^{b}Haplotype AAGGA.

Table 8

Parameter estimation results

Model	Mean		Standard deviation	
diplotypic	theoretical	estimate	theoretical	estimate
h_+/h_+	2.00	2.00 ± 0.27	1.41	1.36 ± 0.28
h_+/h_-	5.24	5.80 ± 0.32	3.41	3.75 ± 0.34
h_-/h_-	8.47	7.98 ± 0.36	4.61	4.62 ± 0.39
haplotypic	theoretical	estimate	theoretical	estimate
h_+	4.25	4.65 ± 0.29	3.31	3.66 ± 0.29
h_-	7.49	7.32 ± 0.28	4.54	4.50 ± 0.32
dominant	theoretical	estimate	theoretical	estimate
h_+/h_+; h_+/h_-	4.66	5.12 ± 0.29	3.38	3.74 ± 0.31
h_-/h_-	8.47	7.99 ± 0.36	4.61	4.63 ± 0.39
recessive	theoretical	estimate	theoretical	estimate
h_+/h_+	2.00	2.00 ± 0.27	1.41	1.36 ± 0.28
h_+/h_-; h_-/h_-	6.97	6.96 ± 0.27	4.40	4.39 ± 0.28

Table 9

Type-I error rates for haplotypes of varying frequency

h_i	Haplotypic					Diplotypic				Dominant			Recessive	
	$T^{\mu,V}$	T^μ	$T^{\mu*}$	T^V	$T^{\mu,V}$	T^μ	$T^{\mu*}$	T^V	$T^{\mu,V}$	T^μ	T^V	$T^{\mu,V}$	T^μ	T^V
1	0.050	0.048	0.051	0.048	0.049	0.047	0.057	0.050	0.047	0.048	0.047	0.053	0.052	0.054
2	0.050	0.050	0.051	0.050	0.051	0.049	0.056	0.052	0.048	0.047	0.049	0.054	0.051	0.055
3	0.048	0.049	0.048	0.049	0.052	0.049	0.060	0.054	0.047	0.048	0.048	0.057	0.052	0.059
4	0.043	0.043	0.047	0.045	0.082	0.045	0.168	0.100	0.041	0.043	0.043	0.106	0.050	0.127
5	0.045	0.044	0.048	0.046	0.110	0.044	0.227	0.137	0.050	0.050	0.051	0.135	0.043	0.164

Simulation model: haplotypic contributions, represented by random variables X_i, X_j, combine in a diplotype as $X_i + X_j$, where $X_i \sim N(1, 1)$.

h_i: tested haplotype (see Table 1).

Table 10

Type-I error rates and power for the overall haplotypic tests

Settings	HTR	HTR$_S$	$T_S^{\mu,V}$	$T_S^{\mu^{\square},V}$	T_S^{μ}	$T_S^{\mu\,\square}$	T_S^{V}
S#0	0.054	0.049	0.046	0.051	0.046	0.049	0.047
S#1	0.626	0.775	0.669	0.677	0.770	0.774	0.050
S#2	0.487	0.607	0.970	0.976	0.597	0.675	0.916
S#3	0.367	0.524	0.770	0.766	0.519	0.503	0.534
S#4	0.763	0.705	0.880	0.898	0.676	0.748	0.644
S#5	0.892	0.756	0.925	0.924	0.734	0.734	0.729
S#6	0.904	0.822	0.693	0.708	0.799	0.812	0.055
S#7	0.784	0.749	0.627	0.637	0.744	0.750	0.053
S#8	0.810	0.957	0.915	0.914	0.956	0.956	0.054

Simulation model: haplotypic contributions, represented by random variables X_i, X_j, combine in a diplotype as $X_i + X_j$; where $X_i \sim N(\mu_i, \sigma_i^2)$.

HTR$_S$, $T_S^{\mu,V}$, $T_S^{\mu^*,V}$, T_S^{μ}, $T_S^{\mu^*}$, T_S^{V}: Simes-based overall tests.

Setting S#1 (single effect, common variance): $\mu_1 = 4; \mu_{i\neq1} = 2; \sigma_i = 6 \; \forall \; i$.

Setting S#2 (single effect with a decreased variance): $\mu_1 = 4; \sigma_1 = 5; \mu_{i\neq1} = 2; \sigma_{i\neq1} = 8$.

Setting S#3 (single effect with an increased variance): $\mu_1 = 4; \sigma_1 = 9; \mu_{i\neq1} = 2; \sigma_{i\neq1} = 7$.

Setting S#4 (all μ_i, σ_i are different): $\mu_i = 9, 8,\ldots,2; \sigma_i = 8, 7,\ldots,1$.

Setting S#5 (all μ_i, σ_i are different): $\mu_i = 9, 8,\ldots,2; \sigma_i = 5, 6,\ldots,12$.

Setting S#6 (all μ_i, are different, common σ): $\mu_i = 9, 8,\ldots,2; \sigma_i = 6 \; \forall \; i$.

Setting S#7 (all μ_i, are different, common σ, the five target haplotypes are equally frequent): $\mu_i = 9, 8,\ldots,2; \sigma_i = 6 \; \forall \; i$.

Setting S#8 (FUSION study: two μ_i's, are different from the rest, common σ): $\mu_3 = 8, \mu_5 = 2, \mu_{i\neq3,5} = 5; \sigma_i = 12 \; \forall \; i$.

Table 11

Type-I error rates and power under violation of normality

Settings	Haplotypic					Diplotypic				Dominant		
	$T^{\mu,V}$	T^{μ}	$T^{\mu*}$	T^{V}	$T^{\mu,V}$	T^{μ}	$T^{\mu*}$	T^{V}	$T^{\mu,V}$	T^{μ}	T^{V}	HTR
S#0a	0.041	0.052	0.051	0.039	0.041	0.049	0.061	0.041	0.041	0.050	0.036	0.050
S#0b	0.193	0.050	0.070	0.228	0.266	0.049	0.089	0.322	0.207	0.050	0.251	0.043
S#1	0.954	0.968	0.961	0.154	0.912	0.942	0.940	0.126	0.934	0.945	0.205	0.969
S#2	0.934	0.964	0.964	0.071	0.872	0.928	0.936	0.064	0.872	0.926	0.045	0.964
S#3	0.976	0.230	0.177	0.985	0.958	0.219	0.159	0.968	0.968	0.153	0.980	0.228
S#4	0.945	0.145	0.197	0.967	0.917	0.088	0.198	0.955	0.857	0.151	0.891	0.141

Simulation model: haplotypic contributions, represented by random variables X_i, X_j, combine in a diplotype as $X_i + X_j$; where $X_i \sim \text{Gamma}(\alpha_i, \beta_i)$; $E(X_i) = \alpha_i\beta_i$; $\text{Var}(X_i)=\alpha_i\beta_i^2$; the sample size was 300 diploid individuals.

BC=1: Box-Cox transformation to normality.

Setting S#0a (H_0): $\alpha_i = \beta_i = 1 \ \forall \ i$; BC=1.

Setting S#0b (H_0): $\alpha_i = \beta_i = 1 \ \forall \ i$; BC=0.

Setting S#1: $\alpha_1 = 1$; $\beta_1 = 1.7$; $\alpha_i = \beta_i = 1$; for $i \neq 1$; BC=1.

Setting S#2: $\alpha_1 = 1.5$; $\beta_1 = 1$; $\alpha_i = \beta_i = 1$; for $i \neq 1$; BC=1.

Setting S#3: $\alpha_1 = 1$; $\beta_1 = 3$; $\alpha_i = 3$; $\beta_i = 1$; for $i \neq 1$; BC=1.

Setting S#4: $\alpha_1 = 3$; $\beta_1 = 1$; $\alpha_i = 1$; $\beta_i = 3$; for $i \neq 1$; BC=1.