Statistical power of association using the extreme discordant phenotype design

Ge Zhang[a], Daniel W. Nebert[a], Ranajit Chakraborty[a] and Li Jin[b,a,c]

Background **Selective genotyping has been proven to be an effective design for mapping quantitative trait loci (QTL), either by linkage or by allelic association, wherein the individual trait values can be used as the indices for phenotype selection. It has also been proposed that association studies of dichotomous traits can benefit from such design. When there is no quantitative measurement for phenotype available, cases and/or controls having extreme discordant phenotypes (EDP) can still be selected, based on their exposure status to a drug toxicity or environmental risk factor. The advantage of EDP design is intuitive and it has been successfully used in a number of studies.**

Methods **In this report, we developed a statistical method to calculate the power of EDP methodology, using a mixture model of genotype-specific distributions of a single biallelic susceptibility locus. We also compared the power of three statistical tests commonly used in association studies – including the χ^2 test of allelic frequencies, the χ^2 test of genotype frequencies, and the Armitage trend test. The power of two different EDP designs was evaluated under a range of scenarios.**

Results and conclusion **Our results indicate that the χ^2 test of genotype frequency is a robust, though less powerful, test for single-locus association, and that EDP methodology is a powerful design for genetic association studies – especially those of common diseases caused by quantifiable drug toxicity or environmental risk factors.** ***Pharmacogenetics and Genomics*** **16:401–413**

Pharmacogenetics and Genomics 2006, **16**:401–413

Keywords: extreme discordant phenotype, environmental exposure, selective genotyping, statistical power, association study

[a]Department of Environmental Health and Center for Environmental Genetics (CEG), University of Cincinnati Medical Center, Cincinnati, OH, USA, [b]State Key Laboratory of Genetic Engineering and Center for Anthropological Studies, School of Life Sciences and Morgan-Tan International Center for Life Sciences, Fudan University, Shanghai, China and [c]CAS-MPG Partner Institute of Computational Biology, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai, China.

Correspondence and requests for reprints to Li Jin, PhD, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, China.
Tel: +1 86 21 65642800; fax : +1 86 21 55664388;
e-mail: ljin007@gmail.com

Received 6 June 2005 **Accepted** 6 December 2005

Introduction

In mapping complex disease genes, genetic epidemiologists take advantage of using selected samples of extreme phenotypes versus random samples to increase the power, based on the fact that such designs are more informative about alleles for disease susceptibility or resistance [1,2]. In quantitative genetics, this idea was applied to linkage analysis of quantitative trait loci (QTL) and termed 'selective genotyping' by Lander and Botstein [3]. It was shown that the power to detect linkage can be increased by ascertaining sib pairs through a proband with extreme phenotype (extreme proband, EP) versus random pairs [4,5]; this strategy was successfully used in identifying several QTLs such as reading disability [6], allergic (atopic) asthma [7], and obesity [8]. Other sib-pair selection strategies for non-parametric linkage analysis, based on identity-by-descent (IBD) sharing [9], were also developed – such as extreme discordant sib-pair (EDSP) design [10,11] and its related derivatives [12–14]. Due to difficulties in identifying and ascertaining such sib pairs exhibiting extreme discordant phenotypes, however, only a few studies with extremely large sample size [15–17] have implemented this design.

The idea of selective genotyping is more feasible for association studies of quantitative traits, in which the sampling is not dependent on the uncommon occurrence of extreme phenotype in multiple relatives. Abecasis *et al.* [18] showed selective genotyping can increase the power by several orders, in detecting allelic association in families. Van *et al.* [19] and Schork *et al.* [20] studied the power of single-locus association analysis, with cases and controls sampled from opposite ends of a quantitative trait distribution, based on a simple-mixture model for genotype-specific trait distributions of a biallelic locus; as expected, this sampling strategy can also increase substantially the power of population-based associations. The result has also been extended to multiallelic settings [21]. Xiong *et al.* [22] studied the impact of phenotypic selection on allele and haplotype frequencies, using a comprehensive analytical approach, and suggested that – instead of comparing allele frequencies between cases and controls – multiple regression methods that employ individual phenotype values could provide higher power than a simple χ^2 test.

The strategies involving selective genotyping have largely been confined to quantitative traits, however, in which

1744-6872

selection is based on a quantitatively measurable trait. Although it has been suggested that the discordant sampling can also be achieved in association studies for dichotomous traits (e.g. disease affliction status) by several selective case–control strategies [1], their statistical performance has yet to be dissected in a mathematically rigorous way. In the context of pharmacogenetics, Nebert [23,24] suggested that one might select the most sensitive (patients who receive the lowest dose for the shortest time and exhibit a drug toxicity response) and the most resistant (patients who receive the highest dose for longest times and exhibit no toxicity) phenotype groups, to find genes responsible for the drug toxicity trait, and this method was referred to as 'extreme discordant phenotype' (EDP) design. In contrast to the traditional selective genotyping designs on individual trait values, this method provides a way of selective sampling for dichotomous traits based on individual exposure levels to a specific quantifiable environmental risk factor (e.g. prescribed drug, cigarette smoking, radon or X-ray exposure, excessive alcohol intake).

In this report, we first introduce the theoretical framework of a change in genotype frequency of a biallelic trait locus, with sample selection on a quantitative trait or disease liability, following the methods of [19,20], and then extend our analysis to two different EDP designs of selective sampling – based on environmental exposure levels. We survey three commonly used statistical tests in single-locus association studies. We then evaluate the power of the proposed selection strategies, under a range of conditions. Our results suggest that EDP design increases the power of association studies, especially for common diseases with known quantifiable environmental risk factors.

Methods

Change in genotype frequency with sample selection

Consider a quantitative trait, or a latent liability of a dichotomous trait (following Falconer's liability-and-threshold model of threshold characters [25]), that is measured by phenotype value x. To simplify discussion without loss of generality, the unit of measurement is chosen so that the distribution of x has a mean (μ) = 0 and variance (σ^2) = 1. Assume that a biallelic candidate locus underlying this trait has two alleles, A and a, where a is the allele conferring higher risk, and i denotes the number of alleles a in the genotypes so that $i = 0$, 1, and 2 – corresponding to A/A, A/a and a/a, respectively. The population distribution of x is thus a composite distribution of three genotype-specific phenotype distributions (GSPD), $f(x|i)$, weighted by their corresponding genotype frequencies (P_i) (Fig. 1):

$$f(x) = \sum_i P_i f(x|i)$$

In selective genotyping, cases and controls are selected from the upper and lower ends of the phenotype distribution, with thresholds T_S and T_C, respectively, corresponding to the desired tail area: $K_S = P(x > T_S)$ for cases, and $K_C = P(x < T_C)$ for controls. The expected genotype frequencies in cases s_i and controls c_i can then be simply expressed as:

$$\begin{cases} s_i = P_i \frac{1-CDF_i(T_s)}{1-CDF(T_s)} \\ c_i = P_i \frac{CDF_i(T_c)}{CDF(T_c)} \end{cases} \quad (1)$$

in which CDF_i and CDF are the cumulative distribution functions for the GSPD $f(x|i)$ and for the population phenotype distribution $f(x)$, respectively.

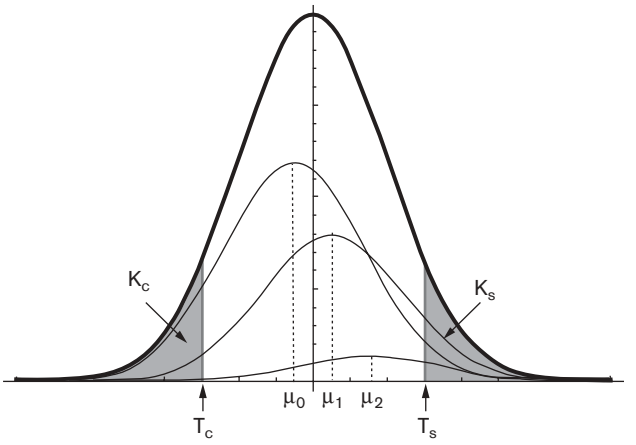
Under the assumption that the contribution of the locus of interest to the phenotype is independent from other factors (i.e. no gene–gene or gene–environmental interactions, which will hereafter be referred to as 'assumption-of-independence'), three GSPDs have the same variance, i.e. $1 - V_G$, where V_G is the locus-specific heritability. We therefore have:

$$E[x|i] = \mu_i$$
$$Var[x|i] = 1 - V_G$$

$$\begin{cases} \sum_i P_i \mu_i = \mu = 0 \\ V_G = \sum_i P_i \mu_i^2 \end{cases} \quad (2)$$

From Equation 2, the three genotypic means μ_i can be obtained, given the locus-specific heritability V_G and the genotype frequencies P_i, under different genetic models.

Fig. 1



Population distribution of the phenotype value x (thick line) and three genotype-specific distributions (thinner lines), weighted by their genotype frequencies. Truncation selections are made at the upper (T_S) and lower (T_C) ends of the distribution, and the shaded areas correspond to the selection areas for cases (K_S) and controls (K_C).

We define the ratio

$$d = \frac{\mu_1 - (\mu_0 + \mu_2)/2}{(\mu_2 - \mu_0)/2}$$

as a measure for different genetic models. This ratio is equivalent to (dominance effect/additive effect). If there is no dominance, $d = 0$. If a is dominant over A, d is positive, and if a is recessive over A, d is negative.

Assuming that each GSPD follows a normal distribution, the genotype frequencies of a specific locus – with known contribution (V_G) in both cases, and controls selected from both ends of the phenotype distribution (with given upper- and lower-tail selection areas) – can be calculated following Equation 1 by replacing CDF_i with the corresponding cumulative normal distribution functions:

$$CDF_i(x) = \Phi\left(\frac{x - \mu_i}{\sqrt{1 - V_G}}\right).$$

Selection by environmental exposure

For a dichotomous trait, EDP design can still be used by selecting samples with different levels of environmental exposure. For example, the most likely cases were selected from none or lightly-exposed affected individuals, whereas the most resistant controls were selected from (heavily) exposed but unaffected individuals. Two different EDP designs are described below.

EDP Design 1

This design applies to the studies in which the exposure is qualitative or categorical (e.g. smoking/non-smoking or exposed/unexposed) and a dichotomous trait is observable. The cases are either none or lightly-exposed affected individuals, whereas the controls are heavily exposed unaffected individuals. First, consider a simple scenario in which the total population is divided into two groups, according to the exposure status to an environmental risk factor with an exposure rate P_{exp} and relative risk r. Then the population distribution of the latent liability x is the weighted sum of the two group-specific distributions: $f(x|\text{exp})$ for the exposed group and $f(x|\text{non})$ for the non-exposed group (Fig. 2a).

$$f(x) = P_{exp} f(x|\text{exp}) + (1 - P_{exp}) f(x|\text{non})$$

And the relative risk r is:

$$r = \frac{1 - CDF_{exp}(T)}{1 - CDF_{non}(T)} \qquad (3)$$

in which CDF_{exp} and CDF_{non} are cumulative distribution functions for $f(x|\text{exp})$ and $f(x|\text{non})$, respectively, and T is the specified selection threshold corresponding to the disease prevalence K.

If the exposure effect is assumed to be independent from other variables, then the two group-specific distributions $f(x|\text{exp})$ and $f(x|\text{non})$ have the same variance, $1 - V_E$, where V_E is the phenotype variance attributable to the environmental exposure. Similar to Equation 2, the means of the two distributions (μ_{exp} and μ_{non}) and V_E follow:

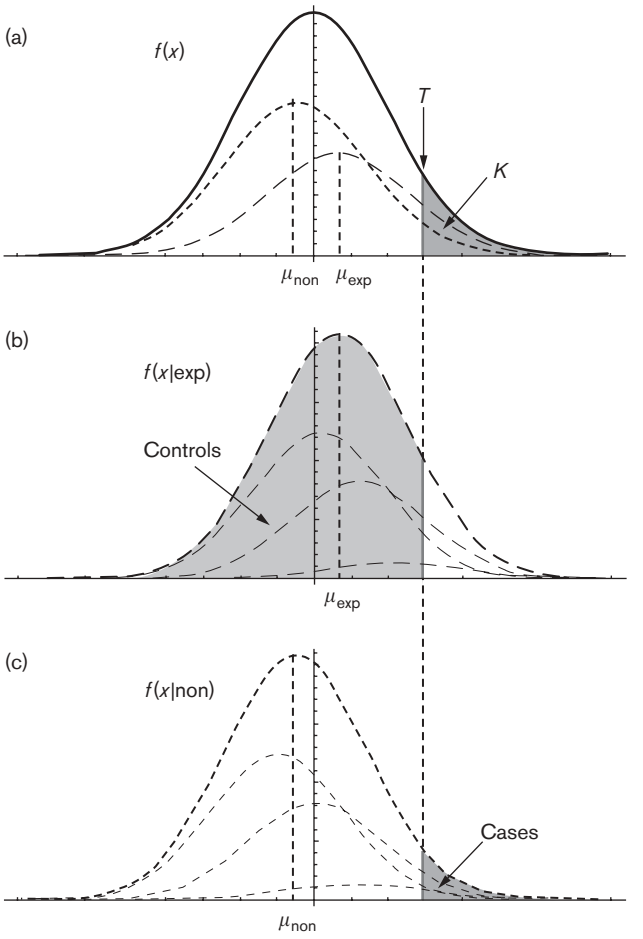
$$\begin{cases} \mu = P_{exp}\mu_{exp} + (1 - P_{exp})\mu_{non} = 0 \\ V_E = P_{exp}\mu_{exp}^2 + (1 - P_{exp})\mu_{non}^2 \end{cases} \qquad (4)$$

Under the assumption of normality, μ_{exp} and μ_{non} can be solved numerically from Equations 3 and 4, given the exposure parameters P_{exp} and r.

The group-specific distributions can be dissected further as the weighted combination of the three GSPDs within each group (Fig. 2b, c):

$$\begin{cases} f(x|\text{exp}) = \sum_i P_i f(x|\text{exp}, i) \\ f(x|\text{non}) = \sum_i P_i f(x|\text{non}, i) \end{cases}$$

Fig. 2



(a) Population distribution of the phenotype value x (thick line) can be expressed as the sum of the phenotype distributions of the exposed group (dashed line) and the non-exposed group (dotted line), weighted by their probabilities. The phenotype distributions of the exposed group (b) and the non-exposed group (c) can be further dissected as the mixture of genotype-specific distributions weighted by their genotype frequencies. Controls are selected from the exposed group (b), and cases are selected from the non-exposed group (c).

where $f(x|\text{exp}, i)$ and $f(x|\text{non}, i)$ are the three GSPDs for the exposed group and the non-exposed group, respectively.

Similar to Equation 1, the expected genotype frequencies for cases selected exclusively from the non-exposed group, and controls from the exposed group, can be calculated as:

$$\begin{cases} s_i = P_i \frac{1-CDF_{\text{non},i}(T)}{1-CDF_{\text{non}}(T)} \\ c_i = P_i \frac{CDF_{\text{exp},i}(T)}{CDF_{\text{exp}}(T)} \end{cases} \quad (5)$$

in which $CDF_{\text{exp},i}$ and $CDF_{\text{non},i}$ are the cumulative distribution functions for the corresponding GSPD functions $f(x|\text{exp}, i)$ and $f(x|\text{non}, i)$, respectively. Note that CDF_{exp} and CDF_{non} are the cumulative phenotype distribution functions for the two groups.

Under the assumption of independence, all of the GSPDs within the exposed or non-exposed group have the same variance, $1-V_E-V_G$, and the means for these distributions are μ_{exp} and μ_{non} shifted by μ_i:

$$Var[x|\text{non}, i] = Var[x|\text{exp}, i] = 1 - V_E - V_G$$
$$E[x|\text{non}, i] = \mu_{\text{non}} + \mu_i$$
$$E[x|\text{exp}, i] = \mu_{\text{exp}} + \mu_i$$

Therefore, under the assumption of normality, the GSPDs for the exposed and non-exposed groups are respectively $N(\mu_{\text{exp}} + \mu_i, 1 - V_E - V_G)$ and $N(\mu_{\text{non}} + \mu_i, 1 - V_E - V_G)$.

Therefore, the genotype frequencies of a specific locus, with a known amount of contribution (V_G) in both cases and controls, selected from the exposed group and the non-exposed group, can be calculated by following Equation 5.

EDP Design 2

This design is applicable to the studies in which the amount of exposure can be measured quantitatively (e.g. known dosage of a drug, cigarette pack-years), and a dichotomous trait is observable. If the amount of exposure e to an environmental risk factor follows a continuous distribution $h(e)$, then instead of selecting cases and controls based on their phenotype value x (which is unobservable), the EDP design can still be achieved by selecting cases and controls from the lower and upper ends of the exposure distribution, with thresholds e_l and e_u corresponding to certain given selection areas K_l and K_u. In other words, cases are selected from affected individuals whose exposure is lower than certain threshold e_l, and controls are selected from unaffected individuals whose exposure is higher than e_u. To simplify the derivation, we assume that the phenotype value is a linear function of the amount of exposure, and we therefore have:

$$\mu_e = E[x|e] = ke + c$$
$$E[x] = E[E[x|e]] = \int \mu_e h(e)\,\mathrm{d}e = \mu = 0$$
$$V_E = Var[E[x|e]] = \int \mu_e^2 h(e)\,\mathrm{d}e$$

where μ_e is the expected phenotype value of individuals with the amount of exposure equal to e, and V_E is the phenotype variance attributable to the environmental exposure. For convenience, we use the transformed variable

$$e^* = (e - E[e])\frac{\sqrt{V_E}}{\sqrt{Var[e]}}$$

to measure the amount of exposure, so that e^* has the density function $h^*(e^*)$ with a mean of zero and variance V_E.

We denote the conditional distribution of the latent liability x under exposure e^* as $f(x|e^*)$, which can also be expressed as the weighted sum of the three GSPDs,

$$f(x|e^*) = \sum_i P_i f(x|e^*, i)$$

in which $f(x|e^*,i)$ denotes the three conditional GSPD functions under exposure e^*.

Under the assumption of independence, the means and variance of the GSPDs, conditional on exposure e^*, are:

$$Var[x|e^*, i] = 1 - V_E - V_G$$
$$E[x|e^*, i] = \mu_{e^*} + \mu_i$$

Where μ_{e^*} is the expected phenotype value, given an exposure equal to e^*. Therefore, the expected genotype frequencies of selected cases, with an amount of exposure lower than e_l, and controls with an amount of exposure higher than e_u, follow the equations:

$$\begin{cases} s_i = P_i \frac{\int_{e^*=-\infty}^{e_l^*} \left(1-CDF_{e^*,i}(T)\right) h^*(e^*)\,\mathrm{d}e^*}{\int_{e^*=-\infty}^{e_l^*} \left(1-CDF_{e^*}(T)\right) h^*(e^*)\,\mathrm{d}e^*} \\ c_i = P_i \frac{\int_{e^*=e_u^*}^{\infty} CDF_{e^*,i}(T)\, h^*(e^*)\,\mathrm{d}e^*}{\int_{e^*=e_u^*}^{\infty} CDF_{e^*}(T)\, h^*(e^*)\,\mathrm{d}e^*} \end{cases} \quad (6)$$

where $CDF_{e^*,i}$ are the cumulative distribution functions for the corresponding conditional GSPD functions $f(x|e^*, i)$ under exposure e^*, and CDF_{e^*} is the cumulative phenotype distribution function for $f(x|e^*)$.

It should be noted that the assumption of normality of the exposure distribution is not necessary. When the amount of exposure (e) within the population follows a particular type of distribution (e.g. normal distribution or

uniform distribution, etc.), then $h^*(e^*)$ follows a scaled distribution of the same type, with the mean equal to zero and variance equal to V_E. Then, given the selection areas of the upper (K_u) and lower (K_l) tails of the exposure distribution, the genotype frequencies in selected cases (from the lower end) and controls (from the upper end) can be calculated by Equation 6.

Power and sample size calculation of association tests

With the expected genotype frequencies available, the power of χ^2 tests on a 2×2 allele frequency table and on a 2×3 genotype frequency table can be easily calculated following standard procedures [26]. The power and sample-size calculation for the Armitage trend test on a 2×3 genotype frequency table is based on the procedure introduced by Freidlin *et al.* [27]. Please refer to Appendices A and B for detail.

In practice, in order to calculate the expected genotype frequencies in both the cases (s_i) and the controls (c_i), one should specify the disease prevalence (K), the genotype frequencies (P_i), the hypothetical contribution (V_G), and the genetic model (defined by ratio d) of the risk locus. For EDP Design 1 and Design 2, one also needs to specify the parameters of the environmental risk factor, based on which cases, and which controls with the extreme discordant phenotype, are selected. In other words, what is the exposure rate (P_{exp}) and relative risk (r) of the environmental risk factor in EDP Design 1; and what is the distribution $h(e)$ and the variance (V_E) attributable to the environmental exposure, and also what are the selection areas K_l and K_u for the cases and the controls in EDP Design 2.

Results

Two-tail vs. one-tail selection

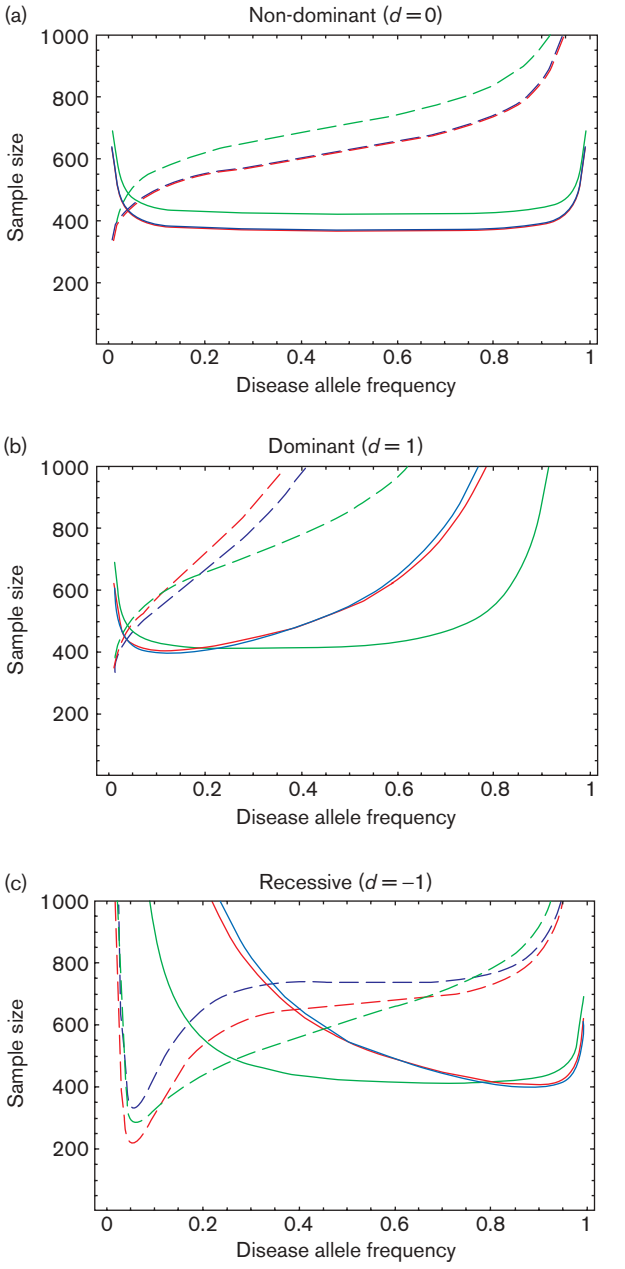
First, we compare the two phenotype-based sample-selection strategies: (1) Two-tail selection is achieved by selecting cases from the upper tail, and controls from the lower tail of the phenotype distribution. For example, one might select the top 5% vs. the bottom 5% extremes for drug-induced hepatic toxicity, as measured by plasma alanine-aminotransferase levels [28], or drug-induced rhabdomyolysis as measured by plasma aspartate-amino-transferase [28], where the selection areas K_S (cases) and K_C (controls) are usually assumed to be equal. Note that this strategy is only valid when the phenotype value of each individual can be measured quantitatively. (2) One-tail selection involves selecting cases from the upper tail of the phenotype distribution and controls from the remaining portion of the distribution ($K_C = 1 - K_S$). Traditional case–control design for dichotomous traits (e.g. affected vs. non-affected or responders vs. non-responders) is therefore an example of this strategy, based on the liability-and-threshold model [25]. The selection area for a case (K_S) is the population prevalence of the disease.

Figure 3 shows the estimated sample sizes required to detect a locus that accounts for 2% of the total phenotype variance ($V_G = 0.02$), with 80% power by two-tail or one-tail selection strategies. For the two-tail selection, the selection areas for both tails are set to $K_S = K_C = 10\%$. In the one-tail selection, the upper-tail area $K_S = 1 - K_C$ is set to 1%, corresponding to a disease with 1% prevalence in the population. Other parameters used for the calculation are listed in the legend to Fig. 3. The results in Fig. 3 indicate that, even with a much looser selection threshold (10% vs. 1%), the two-tail selection is generally more powerful than the one-tail selection – except for a disease allele having a low frequency and under the recessive model. The power of the two-tail selection strategy is less sensitive to the disease allele frequency, even under different genetic models–unless the disease allele has a high frequency (> 0.8) and is dominant, or has a low frequency (< 0.2) and is recessive. In Fig. 3, we also show the power of the three different statistical tests for association. Generally, the power of the χ^2 test of allele frequencies and the power of the Armitage trend test are similar, and both are higher than that of the χ^2 test of genotype frequencies – especially under the non-dominant model. The χ^2 test of genotype frequencies is more robust, however, for determining the disease allele frequency under different genetic models (dominant or recessive), and is substantially more powerful in detecting an association of the high-frequency disease allele, under the dominant model, or the low-frequency disease allele, under the recessive model.

Comparison of the three different statistical tests

We next compared the power of the three different statistical tests under a range of situations. Theoretically, the Armitage trend test and the χ^2 test of allele frequency should be asymptotically equal – if cases and controls are sampled from a population in Hardy–Weinberg equilibrium. These two tests are more powerful than the χ^2 test of genotype frequency, if the allele effect is codominant [29]. We compared these three tests, when the population is under different levels of deviation from Hardy–Weinberg equilibrium (data not shown). The results indicate that the two genotype-based tests (i.e. the trend test and χ^2 test of genotype frequency) are more robust to deviations from Hardy–Weinberg equilibrium; this is because, statistically, F_{st} is a measure of the correlation between the two alleles and, when the population is biased from Hardy–Weinberg equilibrium ($F_{st} \neq 0$), the two alleles are no longer independent. Under this condition, the tally of alleles in cases and controls no longer follows the binomial distribution, which will make the χ^2 test an invalid test, with an increased Type I error (α) when $F_{st} > 0$, or decreased

Fig. 3



Comparison of required sample sizes of the two-tailed selection (solid lines) and one-tailed selection (dashed lines) for detecting association of a biallelic locus with a locus-specific heritability of $V_G = 0.02$. Three different statistical tests are evaluated: the χ^2 test on gene frequencies (red), the Armitage trend test (blue), and the χ^2 test on genotype frequencies (green). The disease allele frequency (*x*-axis) is set from 0.01 to 0.99, assuming that the locus is under Hardy–Weinberg equilibrium within the population. In the two-tailed strategy, the selection areas are $K_S = K_C = 10\%$, and in the one-tailed selection strategy, the upper tail area is $K_S = 1 - K_C = 0.01$. The significance level $\alpha = 0.0001$ and the power $1 - \beta = 80\%$.

Type I error when $F_{st} < 0$. In contrast, the tests of genotype frequency still hold. It should also be noted that the Hardy–Weinberg equilibrium in the total population does not guarantee a Hardy–Weinberg equilibrium in the selected cases and controls under the alternative hypothesis–especially when the risk allele effect is not codominant (i.e. recessive or dominant). Figure 4 shows the comparison of the three tests under different genetic models, with dominance of the disease allele ranging from –1 to +1; here we show that the χ^2 test of genotype frequency, though less powerful, is the most robust test under different genetic models.

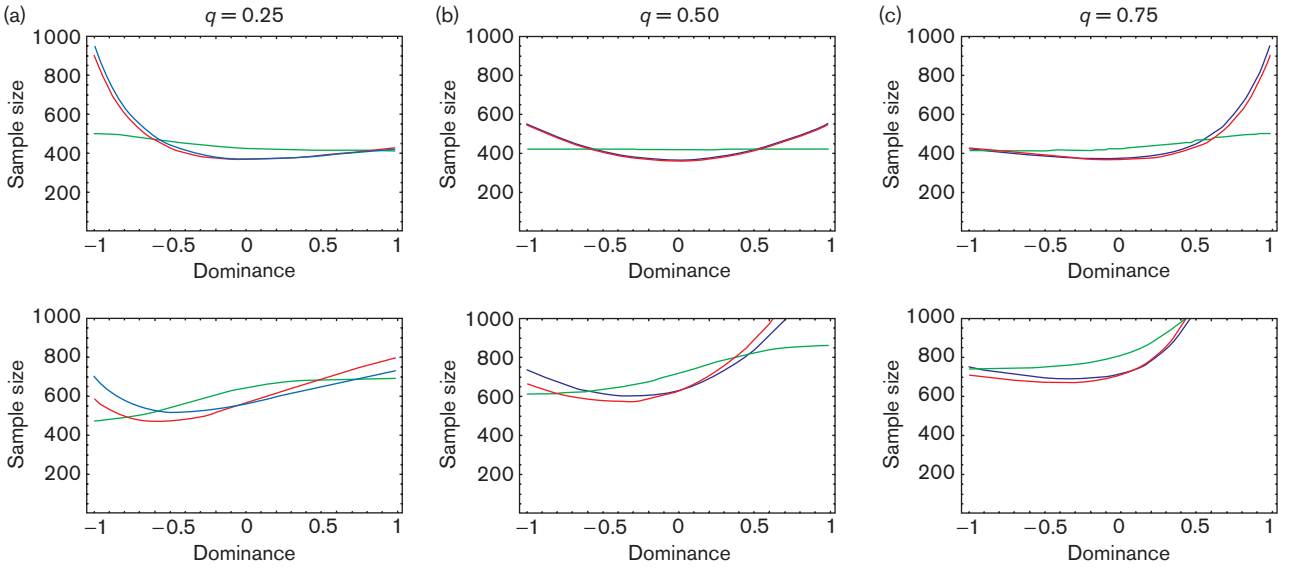
By comparing the power of the two genotype-based tests (Fig. 5), we noticed that the trend test is more powerful than the χ^2 test of genotype frequency, under most conditions (the middle part of the graph between the two black 1 : 1 contour lines). However, the advantage is not substantial, in terms of the sample-size requirement. In some extreme situations, the χ^2 test of genotype frequency can be much more powerful than the trend test. Therefore, we will use the χ^2 test of genotype frequency in the following sections.

EDP design

The results presented above are based on truncation selection strategies (two-tail and one-tail) upon the phenotype value of a quantitative trait, or a latent liability of a dichotomous trait. If there is a known drug or other environmental factor that increases the susceptibility to a disease, the power of the disease association study can be increased by ascertaining cases and controls – conditional on their exposure status. If the exposure is qualitative (exposed versus non-exposed), cases will be selected exclusively from the non-exposed group, and controls from the exposed group (Design 1). If the amount of exposure can be measured quantitatively, then the cases and controls can be drawn from the lower and upper ends of the exposure distribution (Design 2). These designs need to be carefully scrutinized, although they are intuitively valid, given that the shift of the conditional genotypic phenotype distribution under varying levels of exposure will enhance the contrast between cases and controls.

Figure 6 shows the sample size required to detect a biallelic disease locus that accounts for 2% of the total phenotype (or susceptibility) variation with 80% power ($\alpha = 0.0001$), by selecting samples based on different environmental factors having varying relative risk (r). This method has a greater advantage, when the relative risk conferred by the environmental factor is higher. Fig. 6 also shows that the required sample size decreases more quickly with the relative risk, when the disease prevalence is high (Fig. 6b, $K = 0.1$) than when the disease prevalence is low (Fig. 6a, $K = 0.01$), indicating that the method is more effective for common diseases; this turns out to be the case, because a less stringent selection on phenotype (high prevalence) leads to a more

Fig. 4



Comparison of the three tests of association under different genetic models, with estimated required sample sizes for detecting association of a biallelic trait locus with heritability $V_G = 0.02$ by two-tailed selection (upper three graphs) and one-tailed selection (lower three graphs) strategies. Dominance of the disease allele is evaluated from −1 (pure recessive) to +1 (pure dominant). The selection area(s) for two-tailed selection are set at 10%, and for one-tailed selection, 1%. Three different hypothetical disease allele frequencies ($q = 0.25$, 0.50 and 0.75) are used in calculating this figure.

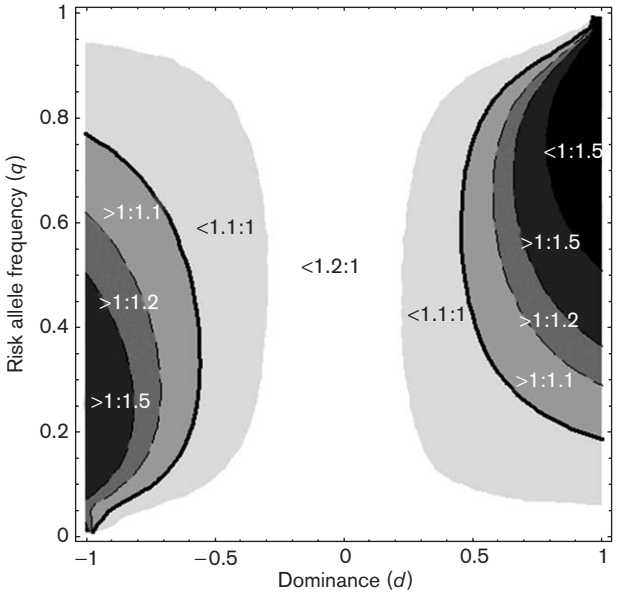
significant shift of the conditional phenotype distributions by controlling for an environmental factor. In Fig. 6, three different exposure rates ($P_{exp} = 0.25,\ 0.5,\ 0.75$) were considered, and they have limited effects on statistical power.

We also compared the power of two different 'partial' conditional sample-selection designs, based on individual exposure status (Fig. 7), in which only cases are exclusively selected from the non-exposed affected individuals, whereas controls are sampled from unaffected individuals – regardless of the exposure status (Fig. 7a), or vice versa (Fig. 7b). The results show striking differences between these two methods: the power of an association study can benefit from using selected cases from non-exposed individuals, but not from using exposed controls, especially when the exposure rate is high.

When the amount of exposure to an environmental risk factor can be measured quantitatively, instead of using the truncation selection upon individual phenotype values, we can still increase the power of an association study by using the truncation selection on individual exposure levels. In contrast to the truncation selection on phenotype values, in this method the cases are selected from affected individuals with low level of exposure (the lower end of the exposure distribution), and the controls are selected from unaffected individuals with high level of exposure (the upper end of the exposure distribution) (Design 2).

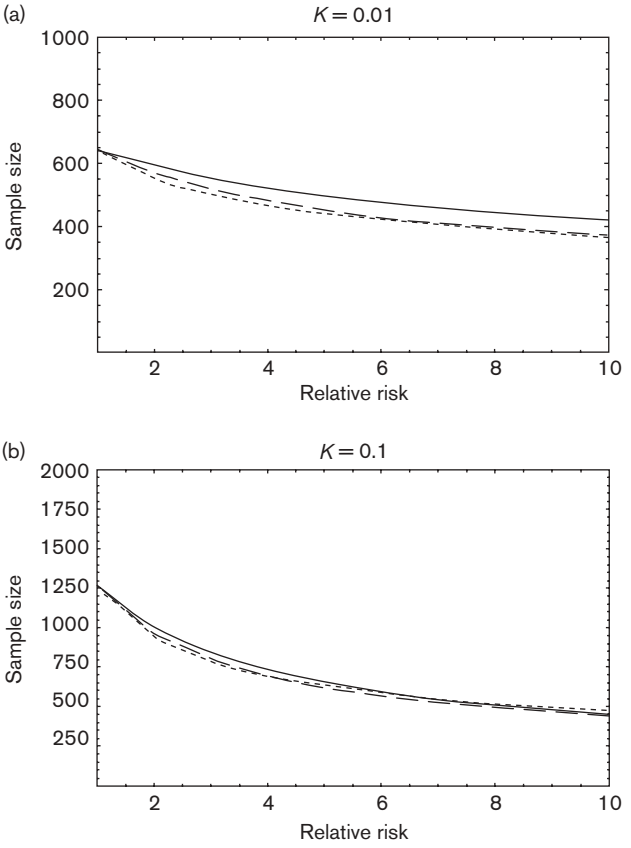
Figure 8 shows that the relationship – between the sample sizes required to detect a biallelic disease locus – accounts for 2% of the total phenotype (or disease susceptibility) variation with 80% power ($\alpha = 0.0001$), by selecting cases and controls based on individual exposure levels to an environmental risk factor with different levels of contribution ($V_E = 0 \sim 30\%$). The percentages of the lower- and upper-tail truncation selection areas of the exposure distribution are also set at three different levels (0.1, 0.2 and 0.5). When the environmental factor is independent of the disease ($V_E = 0$), the sample size is equal to the traditional case–control design without selection. Furthermore, the sample size that is required decreases, when the contribution of the environmental risk factor (V_E) is high. By comparing the gain-of-power of this truncation selection strategy for different disease prevalences (Fig. 8a: $K = 0.01$, and Fig. 8b: $K = 0.1$), it can be seen that the gain-of-power is more when the disease prevalence is high. Figure 8 also shows that the more stringent the selection criteria (i.e. smaller selection area), the higher the gain-of-power by this method. Even under very loose criteria (selection area = 0.5), however, the power can still be substantially increased by using this method, when the environmental risk factor has enough contribution to the disease (e.g. $> 5\%$). In Fig. 9 we studied the relationship between the statistical

Fig. 5



The ratio of estimated sample sizes of the two genotype-frequency tests (χ^2 test : trend test), in order to detect association of a biallelic trait locus with heritability $V_G = 0.02$ under different situations. Between the two black lines (1:1 contour line), the trend test is more powerful than the χ^2 test. The graph is calculated based on the one-tailed selection strategy, which corresponds to the normal case–control design.

Fig. 6



The estimated required sample size to detect association of a biallelic trait locus ($V_G = 0.02$) with EDP design (Design 1), by controlling the environmental risk factor under varying relative risks (1 ~ 10). Different disease prevalences are compared (a: $K = 0.01$; b: $K = 0.1$), and the three different lines (solid, dashed and dotted) represent different exposure rates (0.25, 0.50 and 0.75, respectively).

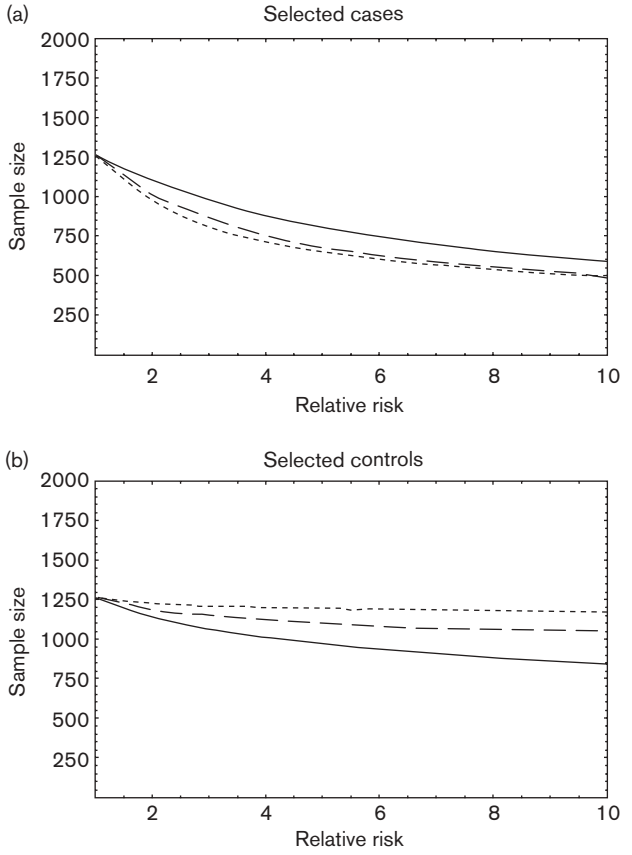
power and levels of selection. The statistical power increases with the stringency of the selection, and this trend is also more obvious in diseases having a high prevalence. Similar to the previous result using Design 1 (Fig. 7), Fig. 10 indicates that the gain-of-power by using this selection method is primarily attributed to the truncation selection cases, below a certain level of environmental exposure. The selection of controls, above a certain level of exposure, has only a minor effect on the statistical power.

During the above discussion of Design 2, we assumed that the amount of exposure is normally distributed, which is rarely the case; moreover, we also required that the amount of exposure has a linear effect on phenotype value. Therefore, we also tested the statistical performance of this selection design under different exposure distributions, e.g. uniform distribution. These results indicated that the statistical power is similar, if the selection criteria are set to be identical.

Discussion

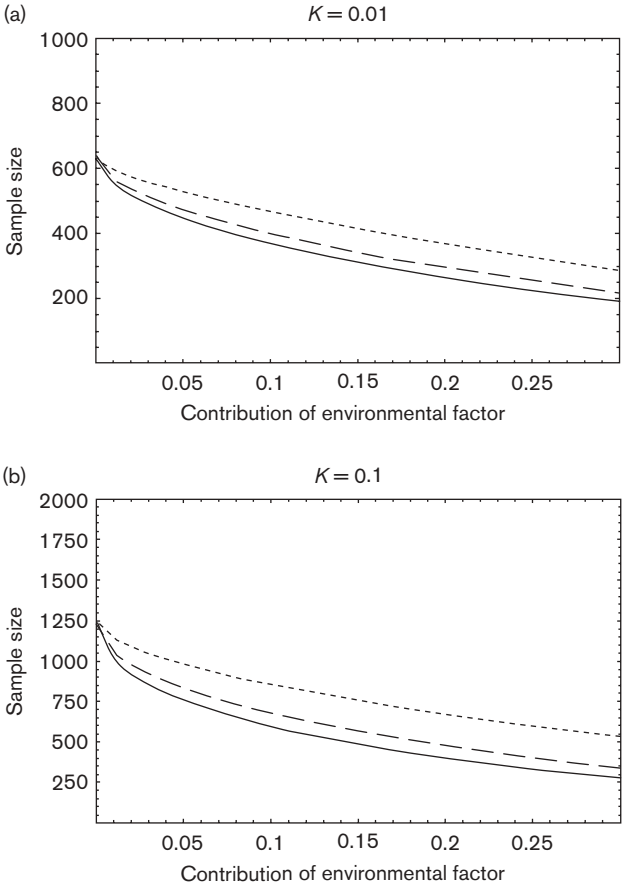
As has been demonstrated previously, selective genotyping can dramatically increase the statistical power in detecting a single-locus association in unrelated individuals. By examining the required sample size of detecting a biallelic locus with a minor contribution (2%) using different sampling strategies (two-tail vs. one-tail), we showed in the present study that the two-tail selection design outperforms the one-tail selection method. More importantly, the two-tail selection with equal upper and lower areas is robust to the disease allele frequency, even when the allelic effect is not additive, and especially by using the χ^2 test of genotype frequency. If the samples are randomly selected from a population and only those individuals who meet the selection criteria are used for genotyping, the two-tail selection strategy is also the most efficient method in terms of phenotyping cost. Furthermore, the optimal proportion for selective genotyping (selection area) can be determined by minimizing the total cost function [21]. We also found that, if the per-individual genotyping cost exceeds the phenotyping cost, the optimal selection area for both cases and controls is around 10%, and if the per-individual genotyping cost is less than the phenotyping cost, then the optimal selection area is around 25% (data not shown).

Fig. 7



The estimated required sample size to detect association of a biallelic trait locus ($V_G = 0.02$) with 'partial' EDP design (Design 1). (a) Only cases are selected from the non-exposed group, whereas controls are randomly sampled from unaffected individuals regardless of their exposure status. (b) Only controls are selectively sampled from the exposed unaffected individuals. The disease prevalence is set at $K = 0.1$, and the three different lines (solid, dashed and dotted) represent different exposure rates (0.25, 0.50 and 0.75, respectively).

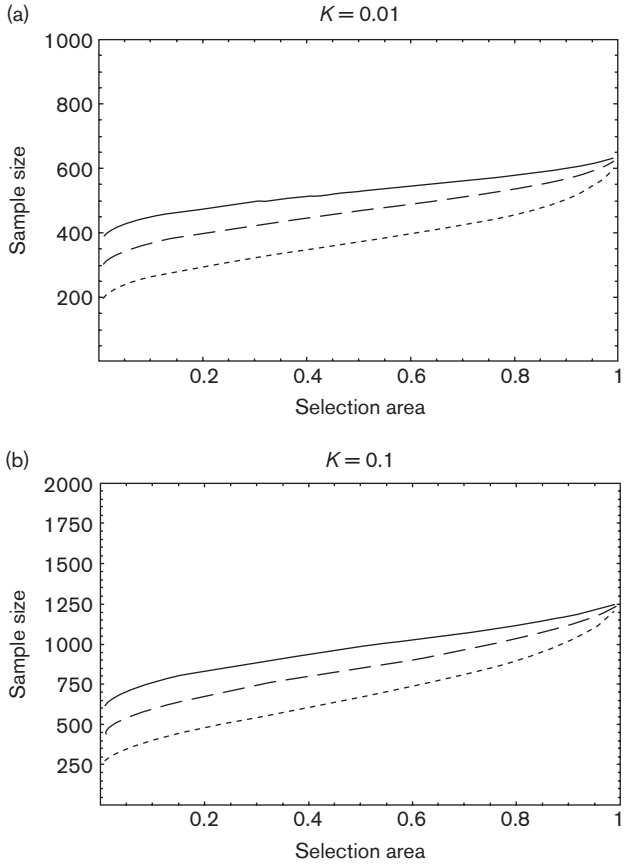
Fig. 8



The estimated required sample size to detect association of a biallelic trait locus ($V_G = 0.02$) with EDP design (Design 2), by controlling the environmental risk factors with varying contributions ($V_E = 0 \sim 30\%$). Different disease prevalences are compared (left: $K = 0.01$; right: $K = 0.1$), and the three different lines (solid, dashed and dotted) represent different selection areas (0.1, 0.2 and 0.5, respectively).

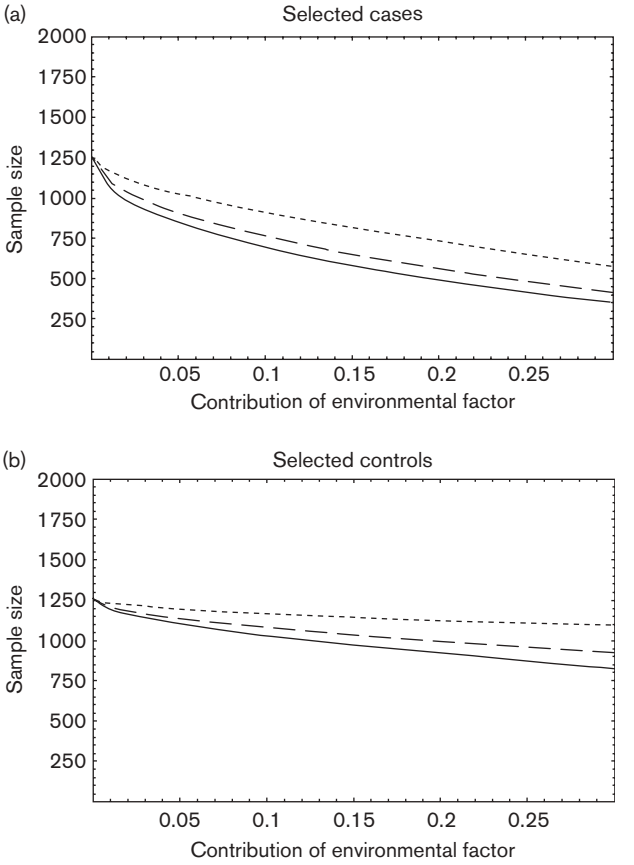
In the present study, we showed that, consistent with the theoretical analysis [29], the χ^2 test of allele frequency is invalid, when the population deviates from Hardy–Weinberg equilibrium, and the χ^2 test of allele frequency will be anticonservative with decreased heterozygosity; this could be another source for a spurious association, at any given Type I error level, even if there is no difference in disease prevalence among the subpopulations. Furthermore, even when the population is in Hardy–Weinberg equilibrium, the genotype frequencies of the trait locus(loci) in cases and controls can still deviate from Hardy–Weinberg equilibrium [30]. This finding is of special relevance to haplotype-based association analyses, in which the current algorithms for haplotype inference depend heavily on the assumption of Hardy–Weinberg equilibrium. The Armitage trend test [31] is recommended by a number of authors [29,32,33] as the genotype-based test for association and also as the best statistical test for the genomic control approach [34,35]; however, our results demonstrate that, although the Armitage trend test is more powerful than the χ^2 test of genotype frequency in most scenarios, the advantage is not substantial. If the disease locus is extremely biased from the additive model, the χ^2 test of genotype frequency can be much more powerful than the trend test; this problem arises from the choice of scores for each genotype in the trend test. Corresponding to the different genetic models, there is a particular set of scores that can be chosen to maximize the power [27,36]. When the genetic model is unknown *a priori*, however, the set of scores corresponding to the additive model (0, 1 and 2 for each genotype) is generally used, which could impair the power when the additive model is violated. A modified trend test has been introduced recently [36], by accessing different sets of scores which provide the maximum and minimum statistical tests; the asymptotic

Fig. 9



The estimated required sample size to detect association of a biallelic trait locus ($V_G = 0.02$) with EDP design (Design 2), by using varying levels of selections (upper and lower selection areas). Different disease prevalences are compared (left: $K = 0.01$; right: $K = 0.1$), and the three different lines (solid, dashed and dotted) represent different contributions of the environmental factors (0.05, 0.1 and 0.2, respectively).

Fig. 10



The estimated required sample size to detect association of a biallelic trait locus ($V_G = 0.02$) with 'partial' EDP design (Design 2). (a) Only cases are selectively sampled. (b) Only controls are selectively sampled. The disease prevalence is set at $K = 0.1$, and the three different lines (solid, dashed and dotted) represent different selection areas (0.1, 0.2 and 0.5, respectively).

distribution for this test is not available, however, meaning that the *P*-value needs to be evaluated by simulation. Therefore, we suggest using the χ^2 test as the genotype-based test instead. In the present study, we also showed the three tests that we compared are not sufficiently powerful in detecting a disease allele having extremely low or high frequencies.

As expected, the EDP design increases the power of the association study for dichotomous traits, by selecting samples based on the level of exposure to an environmental risk factor. Moreover, for both EDP designs discussed in this paper, we have shown that the higher the contribution of the environmental risk factor, the higher the gain of power. In Design 1, the selection is made based simply on the yes/no exposure status to a risk factor, with a certain relative risk and population exposure rate; the influence of sample selection on the power is thus primarily dictated by relative risk. In Design 2, the selection is made based on the amount of exposure to an environmental risk factor having a particular contribution. Besides the contribution of the risk factor, the selection stringency is also related to the power. The gain of power by EDP design is primarily attributable to the selection of cases, however, which implies EDP design with selected cases and random controls is a relatively more economic alternative.

Our results also emphasize the importance of population prevalence of the dichotomous trait under study. The gain of power by using EDP design is much more significant, when the disease prevalence is high. These observations suggest that EDP design can help association studies for common diseases with a known environmental risk factor. Since more and more large cohorts and registry studies for common diseases are becoming

available, and the genotyping cost is still high, due to the large number of markers needed to be typed in genome-wide association studies, EDP design, using selective sampling, provides a practical venue. In particular, the EDP design will be of great value in pharmacogenetics [24], in which the trait (e.g. a toxic response to a certain dose of drug or to a particular environmental exposure such as cigarette pack-years) can be very common and the level of exposure can be quantitatively measured. For example, we are selecting patients with diagnosed head-and-neck squamous cell carcinoma and who have smoked at least one – but not more than 40 – cigarette-pack years (highly sensitive; HS); we are comparing this group to individuals who have no evidence of any type of cancer and who have smoked at least 80 cigarette-pack-years (highly resistant; HR). In another study, we are selecting patients who were diagnosed with gingival overgrowth and who have taken phenytoin not more than 6 months (HS); we are comparing this group to patients who have taken phenytoin for at least 5 years and show no signs of gingival disease (HR). In both clinical studies, candidate genes will be single nucleotide polymorphism (SNP)-typed, comparing the HS and HR groups.

The analytical approach used in this study is based on the mixture model of genotype-specific distribution [19,20], which invokes the implicit assumptions of normality and homoscedasticity (assumption that the variance around the regression line is the same for all values of the predictor) across the three genotypic distributions. The normality assumption can be justified by the nature of complex disease, in which the trait is controlled by numerous factors (genetic or environmental), each of which contributes some limited amount – leading to a gradient from minimally affected to maximally affected. The results presented here were obtained, however, using numerical solutions in which the assumption of normality on the total trait distribution is not required. Therefore, our results are still valid, even if there is a major gene or a major environmental effect. On the other hand, homoscedasticity depends heavily on the assumption of no epistatic effect, which is rarely a biological reality [37]. In order to model the epistatic effects, one needs to know each joint distribution of all the possible combinations of all the interacting factors [38], and this model would be too complicated for the statistical power study. Nevertheless, the influence of gene–gene interactions on our results is limited; this is due to the fact that the genotype frequency change, in selected samples, is mainly determined by the means of the marginal genotypic distributions of the locus under study. However, gene–environment interactions will have a huge impact on the efficacy of EDP design. In addition to change in variance, the means of conditional genotype distribution conditions on an interacting environmental factor will be shifted. Following this argument, we can expect that the power of EDP design will be diminished, if the locus has a positive correlation with the environmental exposure upon which the sample selection is made.

The power analysis presented here was based on a single-locus association under the ideal situation in which the SNP marker under study is the disease locus itself or is in perfect linkage disequilibrium (LD) with the disease locus ($r^2 = 1$). This seldom happens in practical association studies in which markers are only in LD with the disease locus, and the power to detect association is largely confined by the pattern of pair-wise LD between the marker and the disease locus, as well as their allele frequencies [39]. One measure, r^2 of LD between the marker and susceptibility locus (sometimes denoted as Δ^2), provides a good power estimation of association studies. To achieve roughly the same power at the marker locus, the sample size needs to be increased by a factor of $1/r^2$ [40]; therefore, the results on sample-size calculation in this report can be approximately adjusted, based on r^2 [41]. We are also aware that many association studies involve numerous SNPs selected from a candidate gene or genomic segment; consequently, the haplotype-based association tests [42–44] or multi-locus methods [45,46] might be more powerful. The power estimation for such analyses would require prior knowledge about the LD of the markers and can be challenging from a practical standpoint.

Another important issue, which was not addressed in this report, is population stratification; this could be a major source of spurious allelic association [47,48]. Following the genomic control (GC) approach [34,35,49,50], the sample-size and power calculation for an admixed population can be adjusted by introducing a variance-inflation factor (λ), estimated from the empirical distribution of a set of unlinked ('null') loci. It can be demonstrated that a sample-size correction, simply by multiplying the factor λ, provides a rough conservative approximation for an admixed population (data not shown).

Acknowledgements

We thank our colleagues for critical readings of this manuscript and valuable suggestions, and an extremely helpful anonymous reviewer who provided some very important insights that helped us in describing some of our mathematical analyses. Funded in part by NIH P30 ES06096 (D.W.N., R.C. & L.J.).

References

1 Morton NE, Collins A. Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A* 1998; **95**:11389–11393.
2 Vink JM, Boomsma DI. Gene finding strategies. *Biol Psychol* 2002; **61**: 53–71.
3 Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 1989; **121**:185–199.

4 Carey G, Williamson J. Linkage analysis of quantitative traits: increased power by using selected samples. *Am J Hum Genet* 1991; **49**:786–796.
5 Cardon LR, Fulker DW. The power of interval mapping of quantitative trait loci, using selected sib pairs. *Am J Hum Genet* 1994; **55**:825–833.
6 Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. Quantitative trait locus for reading disability on chromosome 6. *Science* 1994; **266**:276–279.
7 Daniels SE, Bhattacharrya S, James A, Leaves NI, Young A, Hill MR, *et al.* A genome-wide search for quantitative trait loci underlying asthma. *Nature* 1996; **383**:247–250.
8 Hager J, Dina C, Francke S, Dubois S, Houari M, Vatin V, *et al.* A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. *Nat Genet* 1998; **20**:304–308.
9 Haseman JK, Elston RC. The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 1972; **2**:3–19.
10 Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 1995; **268**:1584–1589.
11 Risch NJ, Zhang H. Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. *Am J Hum Genet* 1996; **58**:836–843.
12 Gu C, Todorov A, Rao DC. Combining extremely concordant sibpairs with extremely discordant sibpairs provides a cost effective way to linkage analysis of quantitative trait loci. *Genet Epidemiol* 1996; **13**:513–533.
13 Knapp M. Evaluation of a restricted likelihood ratio test for mapping quantitative trait loci with extreme discordant sib pairs. *Ann Hum Genet* 1998; **62** (**Pt 1**):75–87.
14 Rogus JJ, Harrington DP, Jorgenson E, Xu X. Effectiveness of extreme discordant sib pairs to detect oligogenic disease loci. *Genet Epidemiol* 1997; **14**:879–884.
15 Xu X, Rogus JJ, Terwedow HA, Yang J, Wang Z, Chen C, *et al.* An extreme-sib-pair genome scan for genes regulating blood pressure. *Am J Hum Genet* 1999; **64**:1694–1701.
16 Martin N, Goodwin G, Fairburn C, Wilson R, Allison D, Cardon LR, *et al.* A population-based study of personality in 34 000 sib-pairs. *Twin Res* 2000; **3**:310–315.
17 Fullerton J, Cubin M, Tiwari H, Wang C, Bomhra A, Davidson S, *et al.* Linkage analysis of extremely discordant and concordant sibling pairs identifies quantitative-trait loci that influence variation in the human personality trait neuroticism. *Am J Hum Genet* 2003; **72**:879–890.
18 Abecasis GR, Cookson WO, Cardon LR. The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am J Hum Genet* 2001; **68**:1463–1474.
19 Van GS, Houwing-Duistermaat JJ, Adolfsson R, van Duijn CM, Van BC. Power of selective genotyping in genetic association analyses of quantitative traits. *Behav Genet* 2000; **30**:141–146.
20 Schork NJ, Nath SK, Fallin D, Chakravarti A. Linkage disequilibrium analysis of biallelic DNA markers, human quantitative trait loci, and threshold-defined case and control subjects. *Am J Hum Genet* 2000; **67**:1208–1218.
21 Tenesa A, Knott SA, Carothers AD, Visscher PM. Power of linkage disequilibrium mapping to detect a quantitative trait locus (QTL) in selected samples of unrelated individuals. *Ann Hum Genet* 2003; **67**:557–566.
22 Xiong M, Fan R, Jin L. Linkage disequilibrium mapping of quantitative trait loci under truncation selection. *Hum Hered* 2002; **53**:158–172.
23 Nebert DW. Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin Genet* 1999; **56**:247–258.
24 Nebert DW. Extreme discordant phenotype methodology: an intuitive approach to clinical pharmacogenetics. *Eur J Pharmacol* 2000; **410**:107–120.
25 Falconer DS, Mackay TFC. *Introduction to quantitative genetics*. 4th edition. Harlow: Longman; 1996.
26 Agresti A. *Categorical data analysis*. New York: Wiley; 1990.
27 Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case–control studies of genetic markers: power, sample size and robustness. *Hum Hered* 2002; **53**:146–152.
28 Fontana RJ, Lok AS. Noninvasive monitoring of patients with chronic hepatitis C. *Hepatology* 2002; **36**:S57–S64.
29 Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics* 1997; **53**:1253–1261.
30 Knapp M, Seuchter SA, Baur MP. The haplotype-relative-risk (HRR) method for analysis of association in nuclear families. *Am J Hum Genet* 1993; **52**:1085–1093.
31 Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955; **11**:375–386.
32 Slager SL, Schaid DJ. Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. *Hum Hered* 2001; **52**:149–153.
33 Jackson MR, Genin E, Knapp M, Escary JL. Accurate power approximations for chi^2-tests in case–control association studies of complex disease genes. *Ann Hum Genet* 2002; **66**:307–321.
34 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**:997–1004.
35 Bacanu SA, Devlin B, Roeder K. The power of genomic control. *Am J Hum Genet* 2000; **66**:1933–1944.
36 Zheng G. Use of max and min scores for trend tests for association when the genetic model is unknown. *Stat Med* 2003; **22**:2657–2666.
37 Carlborg O, Haley CS. Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* 2004; **5**:618–625.
38 Culverhouse R, Suarez BK, Lin J, Reich T. A perspective on epistasis: limits of models displaying no main effect. *Am J Hum Genet* 2002; **70**:461–471.
39 Muller-Myhsok B, Abel L. Genetic analysis of complex diseases. *Science* 1997; **275**:1328–1329.
40 Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 2001; **69**:1–14.
41 Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004; **74**:106–120.
42 Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *Eur J Hum Genet* 2001; **9**:291–300.
43 Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, *et al.* Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 2001; **293**:489–493.
44 Schaid DJ. Genetic epidemiology and haplotypes. *Genet Epidemiol* 2004; **27**:317–320.
45 Devlin B, Risch N, Roeder K. Disequilibrium mapping: composite likelihood for pairwise disequilibrium. *Genomics* 1996; **36**:1–16.
46 Clayton D, Chapman J, Cooper J. Use of unphased multilocus genotype data in indirect association studies. *Genet Epidemiol* 2004; **27**: 415–428.
47 Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; **361**:598–604.
48 Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet* 2004; **36**:512–517.
49 Devlin B, Roeder K, Wasserman L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* 2001; **60**: 155–166.
50 Reich DE, Goldstein DB. Detecting association in a case–control study while correcting for population stratification. *Genet Epidemiol* 2001; **20**:4–16.

Appendix A

Power and sample-size calculation of χ^2 tests

Under the alternative hypothesis, the χ^2 follows approximately a large-sample non-central χ^2 distribution. Let π_{ij} denote the true probability in cell i, j, and let $\pi_{ij}(M)$ denote the probability to which the maximum likelihood estimate is based on the null hypothesis of no association, where $\sum_{i,j} \pi_{ij} = \sum_{i,j} \pi_{ij}(M) = 1$. For a multinomial sample of size N, the non-centrality parameter for χ^2 is:

$$\lambda = N \sum_{i,j} \frac{\left(\pi_{ij} - \pi_{ij}(M)\right)^2}{\pi_{ij}(M)}$$

Therefore, the asymptotic power for the χ^2 test of association is given by:

$$\text{Power} = 1 - D_{\upsilon,\lambda}(\chi^2_{1-\alpha,\upsilon})$$

in which $D_{\upsilon,\lambda}$is the cumulative distribution of the non-central χ^2 distribution, with υ degrees of freedom, and the non-centrality parameter λ. $\chi^2_{1-\alpha,\upsilon}$ is the $1-\alpha$ lower quantile of the χ^2 distribution, having υ degrees of freedom. The sample size can be calculated by using the above equations in reverse.

Appendix B

Power and sample-size calculation for the Armitage trend test

The power and sample-size calculation for the Armitage trend test on a 2×3 genotype frequency table follows Freidlin *et al.* [27], with minor modifications:

$$\mu_a = \frac{k}{(1+k)^2}\sum_{i=0}^{2} x_i(s_i - c_i)$$

$$\sigma_a^2 = \frac{k^2}{(1+k)^3}\left(\sum_{i=0}^{2} x_i^2 s_i - \left(\sum_{i=0}^{2} x_i c_i\right)^2\right) + \frac{k}{(1+k)^3}\left(\sum_{j=0}^{2} x_i^2 s_i - \left(\sum_{i=0}^{2} x_i c_i\right)^2\right)$$

$$\tilde{\sigma}_a^2 = \frac{k}{(1+k)^3}\left(\sum_{j=0}^{2} x_i^2 s_i - \left(\sum_{i=0}^{2} x_i c_i\right)^2\right) + \frac{k^2}{(1+k)^3}\left(\sum_{i=0}^{2} x_i^2 s_i - \left(\sum_{i=0}^{2} x_i c_i\right)^2\right)$$

$$\text{Power} = \Phi\left(\frac{-z_{1-\alpha/2}(\tilde{\sigma}_a^2 + \mu_a)^{1/2} - N^{1/2}\mu_a}{\sigma_a}\right) + 1 - \Phi\left(\frac{z_{1-\alpha/2}(\tilde{\sigma}_a^2 + \mu_a)^{1/2} - N^{1/2}\mu_a}{\sigma_a}\right)$$

$$\text{Sample Size} = \left(\frac{z_{1-\alpha/2}(\tilde{\sigma}_a^2 + \mu_a)^{1/2} + z_{1-\beta}\sigma_a}{\mu_a}\right)^2$$

in which N is the sample size (cases plus controls) and k is the ratio of the number of controls to the number of cases. s_i and c_i are genotype frequencies in cases and controls. x_i represents the scores used in the trend test, which are usually set as $x_i = i$.