Am. J. Hum. Genet. 68:1463–1474, 2001

The Power to Detect Linkage Disequilibrium with Quantitative Traits in Selected Samples

Gonçalo R. Abecasis, William O. C. Cookson, and Lon R. Cardon

University of Oxford, Oxford, United Kingdom

Results from power studies for linkage detection have led to many ongoing and planned collections of phenotypically extreme nuclear families. Given the great expense of collecting these families and the imminent availability of a dense diallelic marker map, the families are likely to be used in allelic-association as well as linkage studies. However, optimal selection strategies for linkage may not be equally powerful for association. We examine the power to detect linkage disequilibrium for quantitative traits after phenotypic selection. The results encompass six selection strategies that are in widespread use, including single selection (two designs), affected sib pairs, concordant and discordant pairs, and the extreme-concordant and -discordant design. Selection of sibships on the basis of one extreme proband with high or low trait scores provides as much power as discordant sib pairs but requires the screening and phenotyping of substantially fewer initial families from which to select. Analysis of the role of allele frequencies within each selection design indicates that common trait alleles generally offer the most power, but similarities between the marker- and trait-allele frequencies are much more important than the trait-locus frequency alone. Some of the most widespread selection designs, such as single selection, yield power gains only when both the marker and quantitative trait loci (QTL) are relatively rare in the population. In contrast, discordant pairs and the extreme-proband design provide power for the broadest range of QTL–marker-allele frequency differences. Overall, proband selection from either tail provides the best balance of power, robustness, and simplicity of ascertainment for family-based association analysis.

Introduction

Selection of individuals with extreme phenotypes substantially increases power over random sampling for detection of linkage with quantitative trait loci (QTL) (Lander and Botstein 1989; Carey and Williamson 1991; Cardon and Fulker 1994; Eaves and Meyer 1994; Risch and Zhang 1995; Zhang and Risch 1996). For small nuclear families, siblings with discordant phenotypes generally provide the largest gains in power (Risch and Zhang 1995, 1996), but they are difficult to identify and ascertain for genotyping. Other strategies, such as selection of a pair on the basis of only one sibling's phenotype (single selection) or selection of pairs having similar and extreme phenotypes (concordant selection), are easier to implement but provide smaller increases in power. In support of the theoretical predictions, linkage analyses of selected family samples have identified regions that are likely to contain genes controlling diverse quantitative traits (Cardon et al. 1994; Daniels et al. 1996; Hager et al. 1998). More recently, large family and population surveys have been conducted that include 10,000s–100,000s of individuals (Abbott 2000; Martin et al. 2000*b*), and, due to practical constraints on genotyping, these samples are likely to require phenotypic selection.

The influence of selection on the power of model-free linkage analysis has been exhaustively studied. For example, some selection strategies are tailored to the detection of relatively rare trait-increasing alleles, and they influence power by changing expected allele and genotype frequencies from their levels in the general population (Lynch and Walsh 1998). Similar principles may apply to allelic-association analysis (Slatkin 1999), where the focus of interest is on trait-allele relationships at the individual level rather than on allele sharing in families. Although some studies have examined the sampling properties of specific association designs, particularly the popular discordant-sib–pair (DSP) approach (Allison 1997; Allison et al. 1999; Schork et al. 2000), the power to detect allelic association by use of other selection designs has not been systematically explored.

Detection of allelic association with multifactorial traits is the driving force behind the development of fine-scale maps of single-nucleotide polymorphisms (SNPs) across the genome (Chakravarti 1998; Roberts 2000). More than 1.5 million common SNPs are already

Received December 15, 2000; accepted for publication March 28, 2001; electronically published May 8, 2001.

Address for correspondence and reprints: Dr. Lon Cardon, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, United Kingdom. E-mail: lon.cardon@well.ox.ac.uk

0002-9297/2001/6806-0018$02.00

in the public domain (Sachidanandam et al. 2001), providing an average density that may be sufficient to identify complex-trait loci (Collins et al. 1999; Kruglyak 1999; Abecasis et al. 2001). Coupled with the ongoing population-based collections for phenotypic subselection or high-throughput genotyping, the emerging SNP map offers unprecedented opportunities for association analysis.

The largest SNP detection efforts (Altshuler et al. 2000; Mullikin et al. 2000) have focused on detection of relatively frequent polymorphisms, in accordance with the common-disease/common-variant hypothesis (Chakravarti 1998). Analysis of loci for which this hypothesis is correct will require relatively modest sample sizes to identify important mutations (Risch and Merikangas 1996; Risch 2000). However, even in these large SNP maps the disease allele may be unobserved, and, for some loci, there may be multiple trait alleles for which association is not well described by SNPs (Weiss and Terwilliger 2000). In these cases, the issue of allele frequency and selected sampling is less clear, because the power to detect association will be influenced by the frequency of both the marker and the QTL alleles.

Here we examine the power to detect allelic association with QTLs in the selection strategies most commonly used for sib pairs, including single selection, DSPs, concordant sib pairs (CSPs) (in either tail), affected sib pairs (ASPs), and combinations of CSPs and DSPs. We conduct a series of simulations to compare the relative power of the various designs, assuming common marker- and QTL-allele frequencies, and then consider the power to detect association under the full range of marker- and QTL-allele frequencies. In all cases, we use recent QTL-association methods (Abecasis et al. 2000*a*; Monks and Kaplan 2000), which are flexible for analysis of different selection models (and, therefore, trait distributions), subsume nearly all other family-based QTL-association tests (Abecasis et al. 2000*b*), and allow accurate power calculation through a permutation framework.

Methods

QTL Model

We use the standard biometrical model (Falconer 1981), in which the sibling trait, Y_{ij}, for the jth sibling ($j = 1 \ldots n_i$) in the ith family ($i = 1 \ldots N$) is a function of an overall mean (μ), a major additive genetic effect (g_{ij}), residual shared familial effects (s_{ij}), and random environmental effects (e_{ij}), described as

$$Y_{ij} = \mu + g_{ij} + s_{ij} + e_{ij} \ .$$

We assume that the background genetic and environmental effects are normally distributed, with mean zero. For any candidate marker with alleles arbitrarily labeled "1" and "2," let G_{ij} be the number of "1" alleles for the jth offspring in family i. In addition, define the parental genotypes for each family as G_{iM} and G_{iF} (for the male and female parents, respectively). The family-based tests of linkage disequilibrium evaluated here are designed to detect evidence for association between G_{ij} and Y_{ij}, while controlling for the possible effects of population substructure.

Tests of Association

A number of tests have recently been developed for analysis of association in humans (reviewed in Abecasis et al. 2000*b*). The properties of most of these tests are encompassed in the pedigree disequilibrium test (PDT) and related tests (Martin et al. 1997, 2000*a*, 2001; Monks and Kaplan 2000)—which use simple score statistics for association and make almost no assumptions about the phenotypic distribution—and the between- and within-family variance-components model (Fulker et al. 1999; Abecasis et al. 2000*a*), which uses maximum-likelihood approaches to estimate parameters derived from the standard biometrical model. These parameter estimates can be used to help localize the trait locus (Cardon and Abecasis 2000). Both these methods are suitable for general pedigrees, allow for missing parental genotypes, account for population stratification, and are amenable to permutation procedures for non-normal trait data.

The PDT is designed to evaluate the hypothesis that $[Y_{ij} - E(Y_{ij})]w_{ij} \neq 0$, where $E(Y_{ij}) = \bar{Y} = \sum_{ij} Y_{ij} \Big/ \sum_i n_i$ and w_{ij} is a measure of allelic transmission with mean = 0. For nuclear families, a suitable definition for w_{ij} is

$$w_{ij} = \begin{cases} G_{ij} - \dfrac{G_{iF} + G_{iM}}{2} & \text{if parental genotypes are available} \\ G_{ij} - \dfrac{\sum_j G_{ij}}{n_i} & \text{if parental genotypes are not available.} \end{cases}$$

Significance tests are performed via a t test: $T = R/\sqrt{\text{Var}\,(R)}$, where $R = \sum_i R_i$, $R_i = \sum (Y_{ij} - \bar{Y})w_{ij} \Big/ n_i$, and $\text{Var}(R)$ can be estimated as $\text{Var}\,(R) \stackrel{ij}{=} \sum R_i^2$ (Monks and Kaplan 2000). Note that we use the PDT acronym to refer to a family of related tests (Martin et al. 2000*a*; Monks and Kaplan 2000) and that this test is not the discrete-trait test described by Martin et al. (2000*a*).

Association modeling in the variance-components framework seeks to minimize $Y_{ij} - E(Y_{ij})$, using the following expectation of trait scores:

$$E(Y_{ij}) = \mu + \beta_b b_i + \beta_w w_{ij} \ ,$$

where β_b and β_w are orthogonal components of association describing the effect of parental mating type and

allelic transmission, respectively. Using marker genotypes, b_i is defined as

$$b_i = \begin{cases} \dfrac{G_{iF} + G_{iM}}{2} & \text{if parental genotypes are available} \\ \dfrac{\sum_j G_{ij}}{n_i} & \text{if parental genotypes are not available,} \end{cases}$$

and w_{ij} is defined as

$$w_{ij} = G_{ij} - b_i \ .$$

Note that this definition of w_{ij} is equivalent to that shown above for the PDT. Under the null hypothesis of no linkage disequilibrium, $\beta_w = 0$, whereas under the alternative hypothesis, $\beta_w = a$, the additive genetic value at the marker locus (Abecasis et al. 2000*a*).

In the variance components approach, all sources of familial resemblance, including linkage, are modeled separately from association. Evidence for linkage disequilibrium is evaluated by maximizing the likelihood of the data with no constraints on the variance or mean parameters (L_1) and then under the constraint $\beta_w = 0$ (L_0). Under the null hypothesis of no association and when the variance estimates are constrained to be positive (Searle et al. 1992), the quantity $\chi^2 = \ln(L_1) - \ln(L_0)$ is distributed as χ^2 with 1 df (Self and Liang 1987).

To improve efficiency in the present analyses, only families having at least one heterozygous parent (when parental genotypes were used) or at least two different types of offspring (when parental genotypes were ignored) were included in the likelihood calculations, because other families do not contribute to estimates of the β_w parameter. To provide information relevant to study design, all sample sizes presented in the Results section refer to the total number of offspring genotyped, not just the number with heterozygous parents. All analyses were conducted using the QTDT software program (Abecasis et al. 2000*a*).

It is clear that both the PDT and QTDT models derive information from variability within families. Consequently, under some forms of extreme selection (e.g., two siblings from the same tail of the distribution), the loss of phenotypic variability will result in poor performance of these measures. To complement these QTL approaches, we also fitted an extension of the classic discrete-trait transmission/disequilibrium test (TDT) (Spielman et al. 1993) to our simulated data. For these assessments, "affected" and "unaffected" individuals were defined on the basis of a split around the population mean.

We use a variant of the TDT described by Martin et al. (1997) that provides a valid test of linkage disequilibrium, even when multiple affected children per family are considered. In this case, we define $T_i = T_i(1) - T_i(2)$, where $T_i(1)$ and $T_i(2)$ are the total number of transmissions of alleles "1" and "2," respectively, from heterozygous parents to affected offspring in family i. The statistic $T = \sum_i T_i$ is approximately normally distributed with mean zero and a variance of $\sum_i T_i^2$. Therefore, for data sets that include families with more than one child, $t = \sum_i T_i \Big/ \sqrt{\sum_i T_i^2}$ provides a test of disequilibrium.

Selection Strategies

Single-proband and sib-pair ascertainment strategies were defined as having high (Z_H) or low (Z_L) phenotypic thresholds. Although other selection schemes can increase power in linkage studies (Eaves and Meyer 1994; Allison et al. 1999), we focus on those most commonly used in current practice. The following selection schemes were considered: affected proband (AP), in which at least one offspring phenotype exceeds Z_H; extreme proband (EP), in which at least one offspring phenotype is either below Z_L or above Z_H; ASP, in which at least two offspring phenotypes exceed Z_H; CSP, in which ASP or at least two offspring phenotypes are below Z_L; DSP, in which at least one offspring phenotype exceeds Z_H while another is below Z_L; and extreme discordant and concordant (EDAC) sib pair (i.e., CSP or DSP) (Gu et al. 1996).

All schemes were evaluated for selection of families with two or more offspring. The AP and EP schemes were also evaluated for selection of families with a single offspring. Two definitions were considered for Z_H and Z_L. The first was the "truncate" selection: given a desired tail area $\alpha = P(Y > Z_H) = P(Y < Z_L)$, select individuals on the basis of $Z_H = \mu + \Phi_{1-\alpha}^{-1}\sigma$ and $Z_L = \mu + \Phi_{\alpha}^{-1}\sigma$, where μ is the simulated phenotypic mean, σ^2 is the variance, and Φ is the cumulative normal distribution function. The second was "proportional" selection: for a given set of phenotyped families, select Z_H and Z_L so that only a prespecified number are available for analysis, with the constraint that $Z_H - \mu = \mu - Z_L$.

For each selection scheme, we define the selection ratio as the inverse proportion of the sample exceeding the defined threshold, $I = N'/N$, where N' is the number of phenotyped families and N is the number of genotyped families. This measure is designed as an indicator of the difficulty of ascertaining the subsample of interest and is closely related to selection intensity and cost (Falconer 1981 [fig. 11.3 and Appendix A]). The proportional definition of Z_H and Z_L explicitly specifies the selection ratio and is useful for existing random samples, in which it allows comparisons of power for various subsamples of the same size. It is also useful for comparing the attributes of various selection strategies, because, when the ratio of total-to-selected individuals is constant across design, each strategy has the same "difficulty" or cost

of ascertainment. In contrast, truncate selection specifies how extreme probands should be in the trait distribution, regardless of the number of unselected families from which to choose, and is convenient when population-based samples are not available for subselection. In truncate selection, the number of families that must be phenotyped per family selected and the cost of sample collection vary between different schemes and efficiency is less directly comparable.

Simulations

Genotype data were simulated in nuclear families, each having both parents as well as one or more offspring available. Trait offspring values were constructed as the sum of a major gene effect (with variance σ_a^2) generated by a diallelic additive QTL locus, Q, with allele frequencies p_Q and $q_Q = 1 - p_Q$, a residual sibling correlation (σ_s^2), and an environmental effect (σ_e^2), each assigned independently from a normal distribution with mean zero. Unless noted otherwise, a QTL accounting for 10% of the phenotypic variance was simulated, with a background correlation between siblings of .30. For all simulations, a diallelic marker locus "M," with allele frequencies p_M and $q_M = 1 - p_M$, was simulated at a recombination fraction $\theta = .0005$ (under the convention that 1 cM ≈ 1 Mb, this is ~50 kb). For all simulations, we quote the frequency of the trait-increasing QTL and marker alleles.

Linkage disequilibrium between the trait and marker loci was introduced in the parental chromosomes. Disequilibrium was modeled in the usual fashion as $D = p_{MQ} - p_M p_Q$, where p_{MQ} is the frequency of the haplotype with allele "1" at both the marker and trait loci, so that $D_{max} = min(p_M p_Q) - p_M p_Q$, and the standardized disequilibrium coefficient D' is D/D_{max} (Lewontin and Kojima 1960). In assessments of allele frequency differences, unless noted otherwise, D' was held constant under various marker and QTL frequencies.

Empirical Significance Levels and Power

Error rates for these methods are a function not only of the assumptions underlying the particular method but also of trait- and marker-allele frequencies and effect size, since these determine the effective sample size and the shape of the phenotypic distribution. To obtain empirical estimates of power, 1,000 data sets were simulated for each combination of trait- and marker-allele frequencies, degree of disequilibrium, selection strategy, and sample size. From each of these, 50 further sets were generated by randomly permuting the phase of parental chromosomes in each family to generate 50,000 family collections, in which no linkage disequilibrium was present. The 50,000 null-distribution samples were analyzed, and the empirical significance level that gave the desired error rate was estimated. The proportion of the original 1,000 data sets that exceeded this empirical significance level is given as the power of the test under the particular circumstances. In comparison with the empirical-significance approach, asymptotic significance levels were slightly conservative for the discrete-trait TDT (Martin et al. 1997) and the PDT (Monks and Kaplan 2000) but were inflated for the QTDT (Abecasis et al. 2000*a*) after selection. These outcomes are consistent with those observed for variance-component analyses of nonnormal phenotypic distributions (Hopper and Mathews 1982). We use empirical significance levels throughout the Results section, to ensure that the methods are directly comparable and that power does not overly depend on the underlying distributional assumptions of the methods.

Results

First, we consider a relatively common marker (p_m = .4) and a QTL (p_{QTL} = .5) that are separated by a recombination fraction (θ) of .0005 (~50 kb) and in relatively strong disequilibrium ($D' = 0.75$). We then describe the results for variable trait- and marker-allele frequencies.

Table 1 shows the power for designs that include a single offspring and parents. For this family structure, three selection strategies are compared: unselected offspring, AP involving selection on the upper tail of the distribution, and EP involving selection from either tail. The simulations summarized in table 1 employ truncate selection—that is, a fixed number of offspring (360)

Table 1

Power in Families with a Single Offspring, after Selection

Sampling Strategy and Tail Area (Selection Ratio[a])	Power in 360 Single-Child Families when Analysis Method Is		
	TDT	PDT	QTDT
Unselected (1.0)	.15	.54	.55
AP in top tail:			
.05 (20.0)	.86	.02	.03
10 (10.0)	.75	.03	.04
30 (3.3)	.55	.06	.07
50 (2.0)	.36	.11	.12
EP in top or bottom tail:			
.05 (10.0)	.85	1.00	1.00
10 (5.0)	.70	.98	.98
30 (1.7)	.31	.73	.74
50 (1.0)	.13	.54	.53

Note.—Power is the proportion of 1,000 simulated data sets exceeding the 1% empirical significance level, estimated from 50,000 simulations. A diallelic trait locus with equifrequent alleles accounts for 10% of the total trait variance. The marker-locus minor-allele frequency was .40, with θ = .0005 (~50 kb), and $D' = .75$. The residual sibling resemblance is .30.

[a] No. of families phenotyped per family genotyped.

were selected on the basis of trait scores exceeding the prespecified threshold (Z_α or $Z_{1-\alpha}$).

The results in table 1 indicate that both AP selection (i.e., one-tailed selection) and EP selection (i.e., selection from either tail) can provide substantial power increases. Although the maximal power at the .01 significance level is .55 in random samples, selected sampling can increase this to 1.00. Both AP and EP schemes yield considerable gains in power on a per-genotype basis, and, within each strategy, increasing the selection intensity results in increased power. Although the latter trend of increased power with increasing selection is to be expected (Falconer 1981), differences between AP and EP selection highlight an interesting counterexample. Although EP selection requires screening half as many families as AP selection (see table 1), the less-stringent EP strategy has consistently greater power than the more difficult AP design. Thus, for families with one offspring, selection of probands from either tail of the distribution appears most practical and advantageous for statistical power.

Table 1 also shows that maximal power for the AP and EP strategies is not obtained using the same analytical method. The combination of dichotomization of the trait scores and analysis by TDT is clearly not useful for randomly selected samples or samples identified via a proband in either tail (EP), an outcome that is consistent with the well-known advantages of retaining quantitative information rather than arbitrarily dichotomizing continuous measures. In contrast, however, when offspring are collected from only one tail of the phenotypic distribution (AP strategy), there is too little information remaining in the original measurement scale to provide reasonable power. In this case, the TDT is more powerful than the quantitative tests, and even families ascertained under modest selection (e.g., 50%, taking the upper half) should be evaluated in a discrete-trait framework.

Table 2 describes power when sib-pair families are collected. This design allows for mapping genetic effects, using either linkage approaches or linkage disequilibrium tests. The format of table 2 is the same as that of table 1, but it presents the full complement of selection designs to collect samples of 180 sib pairs. In this situation, as expected from previous analyses (Allison 1997; Schork et al. 2000), DSPs provide the most dramatic increases in power, on a per-genotype basis, compared with random selection. For extreme DSPs ascertained from 5% tail areas, the power to detect QTL association is 100%, and it remains >90% for modest selection on 30% tail areas. Aside from DSPs, power is roughly similar for several of the strategies. The relative power of the various approaches may be broadly summarized as DSP > EP ≥ EDAC ≥ CSP > ASP > AP.

Although the DSP strategy is clearly the most powerful for a fixed selected sample size and tail area, ascertainment of DSPs is much more difficult than that of other approaches. Figure 1 shows the ratio of phenotyped-to-genotyped individuals for each of the selection strategies, indicating the relative difficulty of the schemes in table 2. Selection of DSPs requires an initial unselected sample that is >10 times larger than the next-most-difficult scheme (ASP), and the required size of the initial sample is ~300 times larger than that of the easiest scheme (EP). In practice, for the most extreme selection in table 2 ($\alpha = 0.05$), this implies that the 180 DSPs analyzed would have been drawn from >290,000

Table 2

Power in Sib-Pair Families, after Selection

Sampling Strategy and Tail Area (Selection Ratio[a])	Power in 180 Two-Child Families when Analysis Is		
	TDT	PDT	QTDT
Unselected (1.0)	.14	.51	.56
AP with 1 sib in top tail:			
.05 (10.9)	.68	.73	.72
10 (5.7)	.56	.67	.66
30 (2.1)	.29	.55	.54
50 (1.4)	.19	.50	.49
EP with 1 sib in top or bottom tail:			
.05 (5.5)	.39	.92	.93
10 (2.9)	.32	.80	.85
30 (1.2)	.16	.57	.61
50 (1.0)	.12	.50	.56
ASP with 2 sibs in top tail:			
.05 (122.1)	.88	.04	.05
10 (41.5)	.79	.05	.06
30 (7.4)	.40	.11	.12
50 (3.3)	.21	.16	.17
CSP with 2 sibs in the same tail:			
.05 (61.1)	.50	.90	.15
10 (20.8)	.40	.83	.17
30 (3.7)	.17	.54	.29
50 (1.6)	.08	.42	.34
DSP with 1 sib in each tail:			
.05 (1617.9)	1.00	1.00	1.00
10 (217.9)	.99	1.00	1.00
30 (10.1)	.62	.94	.94
50 (2.6)	.25	.67	.65
EDAC with 2 sibs in either tail:			
.05 (58.8)	.53	.94	.50
10 (19.0)	.47	.85	.68
30 (2.7)	.26	.68	.71
50 (1.0)	.13	.46	.51

Note.—Power is the proportion of 1,000 simulated data sets exceeding the 1% empirical significance level, estimated from 50,000 simulations. A diallelic trait locus with equifrequent alleles accounts for 10% of the total trait variance. The marker-locus minor-allele frequency was .40, with $\theta = .0005$ (~50 kb), and $D' = .75$. The residual sibling resemblance is .30.

[a] No. of families phenotyped per family genotyped.

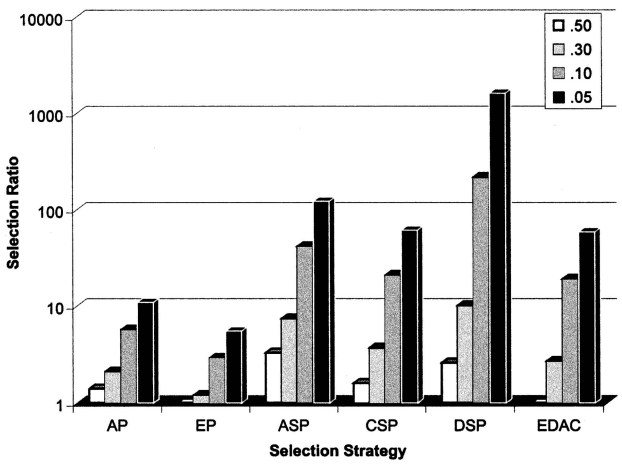

Figure 1 Differences in selection ratio for various sampling designs. Under truncate selection, any threshold of selection (Z_α) requires screening different numbers of families for each selection strategy. The bars show the number of families phenotyped per family selected for genotyping, for $\alpha = .50, .30, .10$, and $.05$.

families in the initial set, whereas the ASPs and EPs would require initial sets of ~22,000 and ~1,000 families, respectively. Thus, DSP selection involves much higher phenotyping costs than those required by any of the other selection schemes. These sampling differences point to the EP method as a practical alternative to DSP selection, because it is the easiest selection scheme yet provides comparable power. Coupled with the outcomes in table 1, these results suggest that selection based on EPs from either tail is efficient and powerful for family-based association analysis of both triad families and small sibships.

The results in table 1 and table 2 are most useful for designing a new study, where there is flexibility in determining the size of the initial sample and the intensity of selection. However, the preceding results are less useful for comparison of subselection strategies in phenotyped cohorts, such as existing family collections (see the description of proportional selection in the Methods section). Table 3 presents power comparisons for the latter situation.

The comparisons in table 3 involve the same number of selected individuals and the same selection ratio across all six selection strategies. The rows show the selection ratio (e.g., 1 is random selection; 20 indicates that the initial sample size is 20 times larger than the selected size). The columns compare the various designs and show power when the trait is considered as discrete (TDT) or continuous (PDT or QTDT).

Again, as expected, DSPs provide considerable increases on a per-genotype basis. However, for a fixed cohort size and selection ratio, DSP is not the most powerful selection strategy: the EP design offers more power than the DSP approach. The difference is greatest at low levels of selection (selection ratio 2–4), where EP selection can yield 30%–50% more power than DSP. These advantages decrease as the intensity increases.

Most of the other selection strategies have similar power at high selection ratios but have consistently less power than EP or DSP selection. Note that the ASP design performs poorly for family-based association (using the TDT), because it induces fixation of common alleles. In general, the pattern of power for the various designs is EP > DSP > EDAC > CSP ⩾ AP > ASP.

Effect of Allele Frequencies

The results presented thus far assume a trait locus with equally frequent alleles and a relatively common SNP marker (minor allele frequency = .40). Although these constraints correspond to the common-disease/common-variant hypothesis and are useful when comparing various selection strategies, they do not permit assessment of the effects of allele frequencies on power. Figure 2 summarizes power for a range of trait- and marker-allele frequencies in unselected samples. Trait-allele frequencies are shown on the *x* axis, and the frequency of the associated marker allele on the *y* axis. Regions of similar power are shown within specific contours. Clearly, the power is greatest when the trait allele is common, because such alleles offer the most within-family genotype variability. Even more important, however, is the frequency difference between trait and marker alleles. Power is maximal whenever the trait- and marker-allele frequencies are identical, and it decreases rapidly when they differ. Thus, for the sample size considered, an optimal trait locus of 50% frequency has >60% power when associated with a marker allele of identical frequency but is virtually undetectable when paired with an allele of 10% frequency, despite the same degree of disequilibrium (D′) in both cases.

The results shown in figure 2 also emphasize the importance of phase for association studies: markers must not only have allele frequencies similar to those of the trait locus, but the specific associated allele must be in phase. For example, a marker-trait pair each having 80:20 allele-frequency patterns can yield dramatic power or, essentially, none at all, depending on whether the QTL allele (of 20% frequency) is associated with the rare (optimal power) or common marker allele (no power). The importance of phase is well known for studies of single-gene disorders but is rarely noted in complex-trait applications.

In selected samples, trait- and marker-allele frequencies also have a major effect on power. Figure 3 shows the relationships of the marker and trait allele frequencies for all the selection strategies considered. In all cases, the allele frequencies refer to the original, unselected

Table 3

Power for Equal Selection Ratios

Analysis Type and Model	Power for Selection Ratio[a] =						
	1 (50)	2 (100)	4 (200)	8 (400)	12 (600)	16 (800)	20 (1,000)
Discrete-trait analysis after dichotomization:							
AP	.05	.12	.21	.26	.27	.33	.39
EP	.05	.12	.15	.16	.19	.18	.19
ASP	.05	.08	.09	.13	.18	.20	.28
CSP	.05	.04	.06	.09	.09	.11	.16
DSP	.05	.08	.18	.27	.31	.28	.40
EDAC	.05	.09	.12	.16	.15	.16	.20
Quantitative-trait analysis:							
AP[b]	.21	.22	.28	.30	.31	.34	.41
EP[b]	.21	.38	.51	.57	.59	.63	.73
ASP	.21	.09	.05	.03	.05	.03	.02
CSP[b]	.21	.17	.26	.30	.31	.32	.48
DSP[b]	.21	.24	.38	.47	.57	.59	.75
EDAC[b]	.21	.31	.36	.43	.42	.44	.49

Note.—Power is the proportion of 1,000 simulated data sets exceeding the 1% empirical significance level, estimated from 50,000 simulations. A diallelic trait locus with equifrequent alleles accounts for 10% of the total trait variance. The marker locus minor allele frequency was .40, with $\theta = .0005$ (~50 kb) and $D' = 0.75$. The residual sibling resemblance is .30. For each selection ratio, 50 sib pairs were selected as described for "proportional" selection. The selection intensity in each column is constant. Under this scheme, the phenotypic tail area varies. Results refer to continuous-trait analysis by PDT or discrete-trait analysis by the TDT.

[a] Nos. in parentheses indicate no. of pairs phenotyped.

[b] Treatment as continuous variable provides more power than does treatment as a discrete variable.

population. For ASPs, the results are shown for the TDT; for all other designs the results were obtained from the PDT or QTDT, whichever provided more power for the specific selection scheme (table 3). All selection strategies can improve power when an appropriate test of association is chosen (e.g., >80% power is possible for specific combinations of trait- and marker-allele frequencies in all selected samples). Nevertheless, there is considerable variability among the designs.

Different selection strategies favor various types of QTL allele profiles. Selection involving only the upper tail of the distribution (e.g., AP and ASP) provides considerable power for rare trait increasing QTL alleles, whose frequencies increase in the genotyped sample. However, if the QTL is common in the population (i.e., the major allele yields increased trait scores), selection on that tail can fix the allele in the genotyped sample and does not enhance power for family-based association tests. Selection designs involving sib pairs from either tail of the distribution (CSP and EDAC) yield roughly uniform power across the range of QTL allele frequencies. Again, the greatest power for all allele frequencies is in DSP and EP designs. For these designs, the power to detect QTL association is >80% for all pairs of well-matched trait- and marker-allele frequencies.

Figure 3 also shows striking differences between selection strategies for the relationship between trait- and marker-allele frequencies. The width and overall shape of the internal contour in each panel of figure 3 indicate the permissible frequency range, that is, the degree to which the marker-allele frequency can deviate from the trait-allele frequency before power is substantially reduced. With the exception of the single-tail selection schemes (AP and ASP), the greatest permissible range for all strategies occurs when the QTL has equifrequent alleles. This is most evident under DSP selection; for example, when the QTL has 50% frequency, a marker of 20%–80% allele frequency still yields >80% power. In contrast, the permissible range for rare (or common) QTL alleles is quite narrow under DSP selection, for example, detection of association with a QTL of 5% frequency has >80% power only for marker alleles having frequencies of 3%–8%. For the single-tail AP and ASP methods, the range is greatest for rare alleles, with AP permitting greater divergence than ASP. The CSP and EDAC designs have more-restricted frequency profiles and retain maximum power only when the marker and trait allele frequencies are fairly similar. EP selection appears uniformly effective for most trait-allele frequencies. The EP design is nearly as permissive as the DSP approach in the middle frequency range and is better than ASPs at the lower range. These results point to EP selection as offering robustness to allele-frequency variation, in addition to providing a useful balance of statistical power and simplicity of ascertainment.

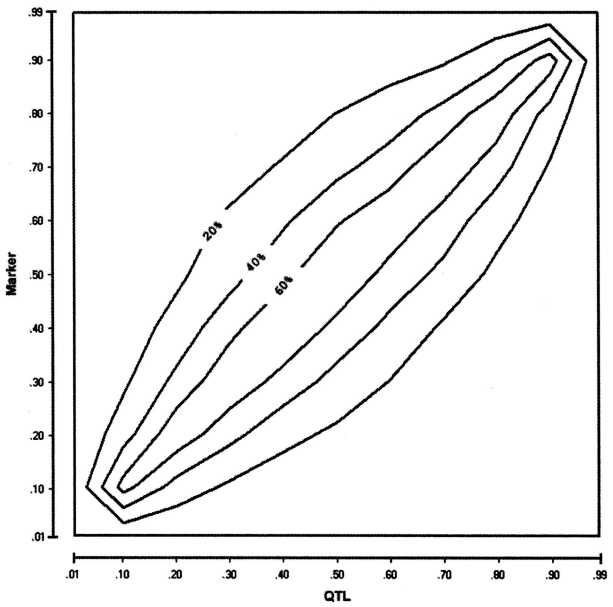

Figure 2 Effect of marker- and trait-allele frequencies on power (random ascertainment). Power was evaluated for 121 combinations of trait- and marker-allele frequencies. In each case, power is the proportion of 1,000 simulated data sets exceeding the empirical 1% significance level, estimated from 50,000 simulations. Each data set included 180 sib pairs where a diallelic trait-locus allele accounts for 10% of the total trait variance. Trait- and marker-locus allele frequencies were as specified, with $\theta = .0005$ (~50 kb), and $D' = 0.75$. The residual sibling resemblance is .30.

Effect of Background Correlation between Siblings

All results presented thus far have assumed a relatively large (30%) background resemblance between siblings, which may correspond to a combination of genetic effects at other loci (≤60%) and shared environmental effects. Although sibling resemblance affects the various tests differently, it does not appear to change their relative ordering. In particular, the efficacy of DSP selection increases as background sibling resemblance increases (10%–50%), whereas the ASP, ESP, and CSP strategies are clearly more powerful for phenotypes with a low degree of background sibling resemblance. The overall shape of power contours is not affected by the degree of background correlation between siblings.

Comparison of Various Tests

Although the main focus of the present study is the evaluation of study-design issues (selection strategies and marker-allele frequencies), it also provides an opportunity to compare some of the available family-based tests of association. We find that both quantitative tests (QTDT and PDT) are generally much more powerful than the TDT whenever genotyped offspring are available in both tails of the distribution.

There are no substantive differences in the performance of the PDT and QTDT in parent-offspring trios, before or after selection (table 1). The QTDT model includes linkage and familial effects as variance components, but their effect size cannot be reliably estimated in selected samples. Although QTDT is more powerful than the PDT in unselected sib pairs (table 2) or in larger families (e.g., in a set of 90 four-offspring families simulated as in table 2, power at the .01 significance level is .55 for QTDT but only .45 for the PDT) this advantage does not hold in strongly selected samples (table 2). When the QTDT is used, we find that care must be taken in the choice of association model for sib pairs. In particular, CSP samples induce a correlation between the number of alleles in each family (the between-family component of association) and allelic transmission (the within-family component of association) so that the two components of association are not orthogonal. That is, in the selected sample, trait-decreasing alleles tend to be transmitted to sibs concordant for low phenotypic scores, as expected, but these tend to occur in families with more trait-decreasing alleles, and the reverse is true for trait-increasing alleles. This reduces power for QTDT in CSP and strongly selected EDAC samples (table 2), but the reduction can be remedied by selecting an association model without the β_b parameter, which is also valid (Fulker et al. 1999). When an appropriate model is selected, we found the QTDT to be more powerful than the PDT for families with three or more offspring.

The ASP design, which removes increasing amounts of variation within the genotyped sample, is poorly suited to quantitative trait-based analyses, particularly at high levels of selection intensity. Nevertheless, even for discrete-trait TDT analysis, ASPs provide less power than other selection strategies with similar selection ratios, emphasizing the gains possible in QTL association analyses of selected samples. Interestingly, if unselected siblings are also available (e.g., if sib trios are selected on the basis of a pair of EPs), power for ASPs is greater with continuous-trait analysis than with TDT modeling (data not shown).

Discussion

As has been demonstrated for linkage studies, selected sampling can dramatically increase power to detect allelic association in families. The present results indicate that careful selection can yield a reduction in genotyping requirements of several orders of magnitude compared with that required for randomly selected samples. As expected on the basis of studies reported elsewhere (Allison et al. 1998; Schork et al. 2000), DSPs provide the most dramatic gains in power for association. However, ascertainment of DSPs requires phenotype screening of many more families than do other strategies, and DSP

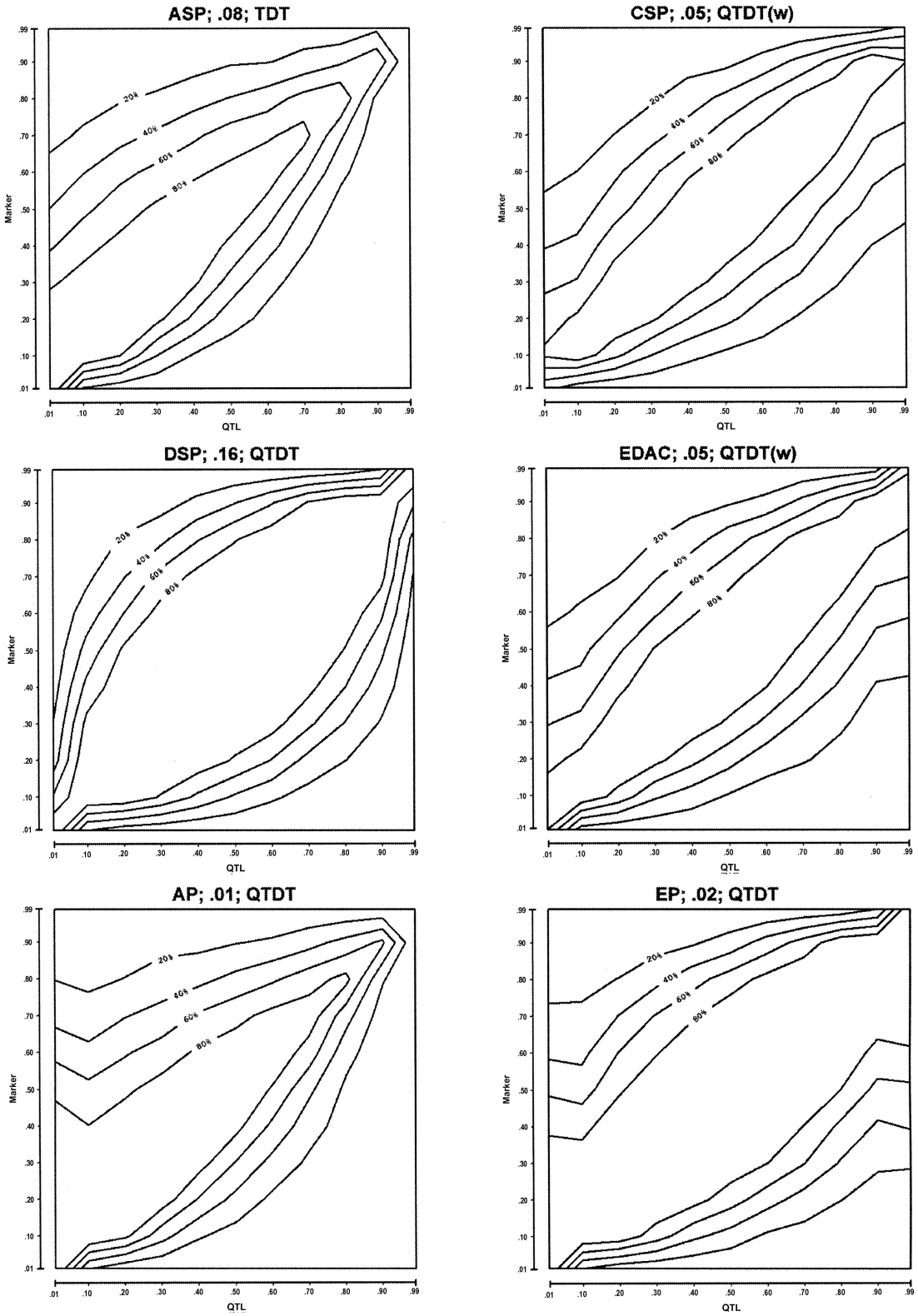

Figure 3 The effect of marker- and trait-allele frequencies on power (under selection). Power was evaluated for 121 combinations of trait- and marker-allele frequencies. In each case, power is the proportion of 1,000 simulated data sets exceeding the empirical 1% significance level, estimated from 50,000 simulations. Each data set included 180 sib-pairs where a diallelic trait-locus allele accounts for 10% of the total trait variance. Trait- and marker-allele frequencies were as specified, with $\theta = .0005$ (~50 kb), and $D' = .75$. The residual sibling resemblance is .30. Thresholds for selection were selected so that ~1 in 50 families was selected for analysis in all strategies ("proportional" selection). The value shown after each label (e.g., ASP; .08) indicates the approximate tail area used to satisfy the selection ratio of 50.

strategies can be difficult to implement. The results reported here suggest that single selection of a proband from either tail of the distribution, coupled with an unselected co-sibling, provides nearly as much power as does the more-stringent DSPs strategy, which requires a 300-fold increase in the number of phenotyped families. Thus, EP selection appears to be a practical, efficient, and powerful sampling design for family-based association studies.

In practice, it is important to consider not only how much power is afforded on a per-genotype basis for each selection strategy but also the cost of sample collection and of particular study designs. If a large registry of phenotyped individuals exists and a subset is to be selected for genotyping, then all selection strategies have the same cost. In this case, we find that selection on the basis of EPs provides similar power to DSPs for very polymorphic trait alleles (~50%) and is more efficient for very rare (<20%) or very common alleles (>80%). If families are to be sampled from the population until a prescribed number meet the selection criteria, the problem is more complex. In this situation, AP or EP collections—which might be started from probands encountered through clinical practice, questionnaire surveys, or other means—are likely to be much less costly than DSP collections, which require prospective phenotyping of many complete families. In addition, for studies involving only a few markers, in which phenotyping costs exceed genotyping costs, it might be reasonable to select a less-efficient selection strategy and absorb the associated increase in genotyping costs.

The results reported here also emphasize the importance of allele frequencies at both the trait and marker loci as suggested in previous studies of discrete traits (Muller-Myhsok and Abel 1997; Tu and Whittemore 1999). Common trait alleles naturally offer more power than rare alleles, but, in general, it is more important to have the marker and trait alleles match in frequency than it is to have only one of them be common (fig. 3). Thus, although the large, ongoing initiatives to identify common SNPs in humans are well suited for detection of QTLs with common variants, it is unlikely that they will perform well with QTLs of low (<20%) or high (>80%) frequency. In these cases, the publicly available SNPs may not be sufficient for association mapping, and additional SNP mining may be required.

There is a clear interaction between selection design and the degree to which marker- and trait-allele frequencies can differ before power decreases. That is, the permissible range of differences in QTL-marker frequency varies as a function of selection strategy. Strategies involving selection on only one tail (ASPs or single AP) tolerate large differences at low frequencies (if the selected tail matches the rarer allele) but are less powerful for common or high-frequency alleles. Alternatively, strategies such as the DSP design allow substantial marker-trait differences in the middle frequency range but only small deviations at the high or low ends of the distribution. The EP-selection model appears most balanced, affording a relatively wide range of permissible frequencies across the entire frequency spectrum.

Availability of parental genotype data also should be considered when choosing a selection strategy. Although all results presented here have made use of parental genotype information, the quantitative tests we examined do not require such information, except in families with only one offspring. Excluding parental genotypes naturally results in a decrease in power for all selected and unselected samples. This decrease is especially significant for strategies that select similar siblings who are likely to be identical at the trait locus (ASP, CSP, or EDAC with a small tail area) so that segregating alleles often cannot be identified. Thus, when parental genotypes are unavailable, these strategies lose virtually all utility and provide little or no power for family-based tests of association. A substantial decrease in power was also observed for the EP and AP strategies, but very little or no power was lost for DSP. In all cases, it is possible to compensate for the reduction in power that results from missing parents by genotyping additional, unselected siblings (Abecasis et al. 2000*a*). However, a word of caution is required, since parental and sibling genotypes help detect pedigree errors, which occur at a low but significant rate in most human samples, and are likely to be enriched in DSPs and other samples of unusual relatives. Selecting a model of association for selected data also requires some forethought about analysis methods, since various models of association for quantitative traits perform better for various selection strategies. For example, QTDT models including both a between- and within-family component association (Fulker et al. 1999; Abecasis et al. 2000*a*) perform poorly when CSPs are analyzed, whereas models including only within-family allelic transmission, such as the PDT, are more powerful in this case (Rabinowitz 1997; Martin et al. 2000*a*). Therefore, in samples consisting mostly of CSPs (i.e., CSP or EDAC), a between-family component of association should not be estimated. In contrast, the between/within QTDT method is more powerful than the PDT as family sizes increase. For example, even a slight increase of sibship size (from two to three) offsets the PDT advantage for CSPs. ASP designs perform best with trait dichotomization and TDT analysis. The power of methods that rely on arbitrary dichotomization of trait values has been thoroughly described elsewhere (Page and Amos 1999).

All our analyses have focused on family-based selection for a single trait. We have considered this trait as resulting from multifactorial influences as well as an

additive QTL of relatively large effect (10%). The present pattern of results does not strongly depend on the assumption of a large QTL, because the same relative ordering of outcomes was observed with a QTL of 5% effect (data not shown). Tests of association can also accommodate dominance and epistatic effects, but the present investigation does not cover these possibilities—for example, mapping dominance and epistatic effects might require even stricter matching of marker- and trait-allele frequencies. The design of association studies is an area of active research, and many interesting issues remain to be explored. In particular, the ability to select for loci of modest effect in the presence of major loci, the ability to explore multiple correlated phenotypes and studies using a mixture of family and unrelated individuals are among the issues deserving further exploration.

The large effect of disequilibrium on power in random samples has been demonstrated by others (Fulker et al. 1999). We have shown that, although selection can further increase power, the detection of association, even after selection, requires that patterns of allele frequencies at the trait and marker loci be similar and that the loci be in phase. In our simulations, common SNP alleles provide only marginal benefits, and they provide none at all if allele frequencies at the QTL are very different. Also, the AP and ASP selection strategies currently in widespread use favor rare alleles and therefore may not be well suited for common diseases. Although DSPs are desirable for common alleles, selection on a single EP and unselected siblings may offer the most practical advantages across the marker- and QTL-frequency range. For any selection strategy, chances of successful localization of human QTLs by association should increase if a variety of SNP markers with different allele frequencies and patterns of disequilibrium are examined.

Acknowledgments

The authors are supported by the Wellcome Trust. This work was also supported in part by grant EY-12562 (to L.R.C.) from the National Eye Institute, National Institutes of Health.

References

Abbott A (2000) Manhattan versus Reykjavik. Nature 406: 340–342

Abecasis GR, Cardon LR, Cookson WO (2000*a*) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–292

Abecasis GR, Cookson WO, Cardon LR (2000*b*) Pedigree tests of transmission disequilibrium. Eur J Hum Genet 8:545–551

Abecasis GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WO (2001) Extent and distribution of linkage disequilibrium in three genomic regions. Am J Hum Genet 68:191–197

Allison DB (1997) Transmission-disequilibrium tests for quantitative traits. Am J Hum Genet 60:676–690

Allison DB, Heo M, Kaplan N, Martin ER (1999) Sibling-based tests of linkage and association for quantitative traits. Am J Hum Genet 64:1754–1763

Allison DB, Heo M, Schork NJ, Wong SL, Elston RC (1998) Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power. Hum Hered 48:97–107

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513–516

Cardon LR, Abecasis GR (2000) Some properties of a variance components model for fine-mapping quantitative trait loci. Behav Genet 30:235–243

Cardon LR, Fulker DW (1994) The power of interval mapping of quantitative trait loci, using selected sib pairs. Am J Hum Genet 55:825–833

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1994) Quantitative trait locus for reading disability on chromosome 6. Science 266:276–279

Carey G, Williamson JA (1991) Linkage analysis of quantitative traits: increased power by using selected samples. Am J Hum Genet 49:786–796

Chakravarti A (1998) It's raining SNPs, hallelujah? Nat Genet 19:216–217

Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 96:15173–15177

Daniels SE, Bhattacharrya S, James A, Leaves NI, Young A, Hill MR, Faux JA, Ryan GF, le Souef PN, Lathrop GM, Musk AW, Cookson WO (1996) A genome-wide search for quantitative trait loci underlying asthma. Nature 383: 247–250

Eaves L, Meyer J (1994) Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. Behav Genet 24:443–455

Falconer DS (1981) Introduction to quantitative genetics. Longman Group, Harlow, United Kingdom

Fulker DW, Cherny SS, Sham PC, Hewitt JK (1999) Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet 64:259–267

Gu C, Todorov A, Rao DC (1996) Combining extremely concordant sibpairs with extremely discordant sibpairs provides a cost effective way to linkage analysis of quantitative trait loci. Genet Epidemiol 13:513–533

Hager J, Dina C, Francke S, Dubois S, Houari M, Vatin V, Vaillant E, Lorentz N, Basdevant A, Clement K, Guy-Grand B, Froguel P (1998) A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. Nat Genet 20:304–308

Hopper JL, Mathews JD (1982) Extensions to multivariate normal models for pedigree analysis. Ann Hum Genet 46: 373–383

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–144

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. Evolution 14:450–472

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA

Martin ER, Bass MP, Kaplan NL (2001) Correcting for a potential bias in the pedigree disequilibrium test. Am J Hum Genet 68:1065–1067

Martin ER, Kaplan NL, Weir BS (1997) Tests for linkage and association in nuclear families. Am J Hum Genet 61: 439–448

Martin ER, Monks SA, Warren LL, Kaplan NL (2000*a*) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67:146–154

Martin N, Goodwin G, Fairburn C, Wilson R, Allison D, Cardon LR, Flint J (2000*b*) A population-based study of personality in 34,000 sib-pairs. Twin Res 3:310–315

Monks SA, Kaplan NL (2000) Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. Am J Hum Genet 66:576–592

Muller-Myhsok B, Abel L (1997) Genetic analysis of complex diseases. Science 275:1328–1329

Mullikin JC, Hunt SE, Cole CG, Mortimore BJ, Rice CM, Burton J, Matthews LH, et al (2000) An SNP map of human chromosome 22. Nature 407:516–520

Page GP, Amos CI (1999) Comparison of linkage-disequilibrium methods for localization of genes influencing quantitative traits in humans. Am J Hum Genet 64:1194–1205

Rabinowitz D (1997) A transmission disequilibrium test for quantitative trait loci. Hum Hered 47:342–350

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Risch N, Zhang H (1995) Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science 268: 1584–1589

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847–856

Risch NJ, Zhang H (1996) Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. Am J Hum Genet 58:836–843

Roberts L (2000) Human genome research: SNP mappers confront reality and find it daunting. Science 287:1898–1899

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Schork NJ, Nath SK, Fallin D, Chakravarti A (2000) Linkage disequilibrium analysis of biallelic DNA markers, human quantitative trait loci, and threshold-defined case and control subjects. Am J Hum Genet 67:1208–1218

Searle SR, Casella G, McCulloch CE (1992) Variance components: Wiley series in probability and mathematical statistics. John Wiley & Sons, New York

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Slatkin M (1999) Disequilibrium mapping of a quantitative-trait locus in an expanding population. Am J Hum Genet 64:1764–1772

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. Am J Hum Genet 52: 506–516

Tu IP, Whittemore AS (1999) Power of association and linkage tests when the disease alleles are unobserved. Am J Hum Genet 64:641–649

Weiss KM, Terwilliger JD (2000) How many diseases does it take to map a gene with SNPs? Nat Genet 26:151–157

Zhang H, Risch N (1996) Mapping quantitative-trait loci in humans by use of extreme concordant sib pairs: selected sampling by parental phenotypes. Am J Hum Genet 59: 951–957