

Alka Malhotra · Kevin Cromer · Mark F. Leppert
Sandra J. Hasstedt

The power to detect genetic linkage for quantitative traits in the Utah CEPH pedigrees

Received: 26 October 2004 / Accepted: 29 November 2004 / Published online: 29 January 2005
© The Japan Society of Human Genetics and Springer-Verlag 2005

Abstract Quantitative trait phenotypes and linked marker genotypes were simulated for a range of models with different sets of assumptions based on displacement, prevalence, and heritability of the trait in 30 Utah Centre d'Etude du Polymorphisme Humain (CEPH) families. The gain in power by the addition of 15 families was also estimated by extrapolation. Power was evaluated using both parametric single locus (PSL) models and variance components (VC) methods for two situations: (1) a single marker with 75% heterozygosity and a recombination fraction of 0.05, and (2) a fully informative marker as an approximation to multipoint analysis. When the simulation and analysis models were both dominant with the same prevalence, power $\geq 80\%$ for $\text{lod} > 3$ was estimated when quantitative trait locus variance was $\geq 40\%$ with a displacement of 2.5 or 3. Power was 5–15% lower for recessive models compared to dominant models. With the addition of 15 families, an average increase in power of 17% and 22% was estimated for the dominant and recessive models, respectively. In PSL analyses, power was estimated at $\leq 20\%$ when the dominance was misspecified. This investigation delineates parameter conditions under which this unique sample affords adequate power to detect linkage using both PSL and VC methods.

Keywords Utah CEPH · Power analysis · Simulation · Quantitative traits · Linkage analysis

Introduction

The Centre d'Etude du Polymorphisme Humain (CEPH) genotype database consists of 61 families with large sibships that have served as a reference panel for the construction of human genetic maps by the international scientific community (Dausset et al. 1990; White et al. 1985). As a result of this collaboration, information has accumulated in these families on thousands of genetic markers, including restriction fragment length polymorphisms, minisatellites, and microsatellites (CEPH genotype database, The Utah Marker Development Group 1995).

Utah CEPH pedigrees constitute the majority (47 pedigrees) of the CEPH reference panel. In recent years, family members have been invited to participate in new clinical interviews and sample collection with support from the Keck Foundation. Multiple ongoing collaborative efforts will continue to yield extensive phenotyping of both quantitative and qualitative traits of biochemical, physiological, and clinical relevance. Furthermore, additional genotyping of a large set of highly informative markers in all Utah CEPH families will soon reach completion, providing a refined map and an extensive database of genotypes on this set of families. This extensive genotyping and phenotyping should make the Utah CEPH families a unique and invaluable resource for analysis of the genetic diversity that underlies individual variation.

The pedigree structures sampled in Utah consisted of a large sibship with at least eight children, their parents, and their living grandparents at the time of sampling. These structures were selected to optimize the probability of identifying genetic linkage: (1) the large sibship size affords replication of segregating events in informative families, and (2) the inclusion of grandparents

A. Malhotra · K. Cromer · M. F. Leppert · S. J. Hasstedt
Department of Human Genetics,
Eccles Institute of Human Genetics,
University of Utah Health Sciences Center,
Salt Lake City, UT, USA

Present address: A. Malhotra (✉)
Diabetes and Obesity Research Unit,
Genetic Basis of Human Disease,
Translational Genomics Research Institute,
445 N Fifth Street, Phoenix, AZ 85004, USA
E-mail: amalhotra@tgen.org
Tel.: +1-602-3438749
Fax: +1-602-3438840

can yield information on the phase of the loci under investigation (Ott 1999). Linkage analysis of all these traits will require large amounts of computer resources and analysis time. Furthermore, the likelihood that linkage will be detected for any particular trait will depend on the degree and the nature of genetic influences. Therefore, before undertaking such a task, it would be of interest to identify the statistical properties of a trait that would increase the chance of detecting linkage. This can be accomplished by estimating the power to detect linkage depending on these characteristics of the trait.

The investigation presented here was restricted to the analysis of quantitative traits. It has often been advocated that unraveling the genetics of complex traits such as common diseases would be markedly facilitated by the genetic analysis of intermediate phenotypes, typically physiological or biochemical parameters closer to gene action. These intermediate phenotypes are generally quantitative variables. Quantitative traits have been the staple of genetics in animal and plant breeding (Lynch and Walsh 1997), and their use in human genetics has long been advocated (Elston 1979; Hase-man and Elston 1972). Analytical methods for the genetic analysis of quantitative traits in human genetics have received increased attention in recent years (Blangero et al. 2000).

In the design of an experiment, it is essential for an investigator to evaluate its statistical power, which in the present case is the probability of detecting linkage given that linkage actually exists (Haines and Pericak-Vance 1998; Risch 1997; Sham 1998). Due to the complexity of the likelihood calculation for family data, power is not readily assessed analytically (Boehnke 1986; Ploughman and Boehnke 1989). Therefore, simulations are used; a large number of replicates of data are simulated based on a priori specified conditions followed by linkage analysis of each of these replicates. The power is estimated as the proportion of replicates displaying lod scores above a given threshold (Ott 1999). A number of published power analyses have addressed similar issues (Duggirala et al. 1997; Greenberg et al. 1998; Rijdsdijk et al. 2001; Shugart et al. 2002; Wijnsman and Amos 1997; Williams and Blangero 1999). Wijnsman and Amos (1997) and Williams and Blangero (1999) compared the power to detect linkage in nuclear families versus extended pedigrees; and another group compared two-point (the use of a single marker) with multipoint (the simultaneous use of multiple markers) linkage analysis (Ekstrom 2001). These studies have documented the expected increase in power as a function of pedigree size and the number of markers used in linkage analysis. Such information has proven extremely useful in interpreting discrepancies or insignificant results caused by the lack of power.

The following question was addressed in our study: for what characteristics of a trait measured on the Utah CEPH pedigrees is there enough information to detect linkage of a marker to a quantitative trait locus (QTL) based on the marker information available? This ques-

tion was answered by simulating data using single locus models and then performing linkage analysis on the simulated data using parametric single locus (PSL) or variance components (VC) models.

Subjects and methods

The Utah CEPH pedigrees

The Utah CEPH families consist of 47 pedigrees with large sibship size, their parents, and their living grandparents. Since 1997, members of 45 Utah CEPH pedigrees (two families are not participating) have been measured for a large number of qualitative and quantitative traits as part of the Utah Genetic Reference Project (UGRP). In its initial phase, completed at the time the present study began, the project had involved the participation of 30 of the 45 Utah CEPH families. (The Utah CEPH pedigree individual numbers can be obtained from the author upon request.) There were 418 individuals in the 30 pedigrees ranging in size from nine to 16 individuals. Of these, 312 were phenotyped, including 16 grandparents, 42 parents, and 254 offspring. Power calculations were performed for this sample set. Extrapolations to the entire set of families are presented in the Results and discussion section. All UGRP study subjects gave informed consent under University of Utah IRB approved protocol number 6090–96.

Assumptions pertaining to marker data for the simulations

The large number of markers that has been genotyped in the Utah CEPH pedigrees have allowed us to make the following assumptions for the simulation of marker data: first, we simulated a marker with heterozygosity of 75% and a recombination fraction of 0.05 between marker and QTL; this was done by assuming four alleles with equal frequencies at the marker locus. Second, as an approximation to multipoint linkage analysis, simulations were performed for a fully informative marker generating four different alleles among the parents with the assumption of a recombination fraction of 0.005. We refer subsequently to these two models as two-point and fully informative, respectively.

Simulations

We defined a large number of simulation models, with a wide range of parameters, to include most of the characteristics that vary among quantitative traits. Phenotypic data were simulated under single locus models. Each simulated QTL was assumed to have two alleles in Hardy-Weinberg equilibrium. Phenotypes were simulated varying the following conditions: (1) dominant or

recessive inheritance; (2) displacement (the difference between the two homozygote means in standard deviation units of the trait distribution conditional on major genotype) of either 2, 2.5, or 3 ; (3) residual heritability (the proportion of the within-genotype variance due to genetic background) equal to 0.1, 0.3, or 0.5; and (4) prevalence (the percentage of individuals who are gene carriers) of 10, 20, or 30%. Specifically, prevalence was equal to q^2 for recessive models and $2pq + q^2$ for dominant models. With all combinations of parameter values, 27 models were generated for each mode of inheritance (dominant or recessive). Recombination fractions were assumed to be equal in males and females.

The pedigree structures were fixed, but both phenotypic and genotypic data were simulated in all cases using the Pedigree Analysis Package (PAP) (Hasstedt 2002). The following steps were used to generate the data: (1) using the two-point (trait and marker) genotype frequencies for a given model, genotypes were assigned to the founders in accordance with Hardy-Weinberg proportions based on selecting a random number; (2) in a similar manner, offspring were assigned genotypes conditional on the parental genotypes, the recombination fraction between the marker and trait locus, and Mendelian transmission probabilities; (3) using the assumed means and standard deviations, quantitative

phenotypes were assigned to each individual using the normal density function. Assignment of genotypes and phenotypes to all individuals in the 30 pedigrees completed one replicate. Five hundred replicates were generated for each model.

Linkage analysis and power estimation

Upon completion of the simulations, linkage analysis was performed on 500 replicates for each model using two methods: PSL and VC. In the PSL method, the model underlying a QTL is specified in terms of mean genotypic effects (under dominant or recessive modes of inheritance), total phenotypic variance, and gene frequency. In the VC method, on the other hand, genetic modeling involves partitioning of the variance into genetic and residual environmental components (Ott 1999). Both the PSL and VC analyses utilize maximum likelihood estimation methods.

For PSL analyses, we assumed either dominant or recessive inheritance with a prevalence of 10% and 30% and a displacement of 2.5. This yielded one VC and four PSL analyses for each replicate. The program LINKAGE (Lathrop et al. 1985) was used for the PSL analyses, and GENEHUNTER (Kruglyak et al. 1996; Pratt

Table 1 Power estimates (%) for lod > 3. Simulation model: recombination fraction of $\theta=0.05$, two-point, dominant; analysis model: dominant. *QTL* quantitative trait locus

Simulation					Analysis		
Total heritability	QTL variance	Residual heritability	<i>t</i> (displacement)	Prevalence	Power 10 ^a	Power 30 ^b	Power VC ^c
0.3382	0.2647	0.1	2.0	10	24.4	6.40	6.20
0.4239	0.3599	0.1	2.5	10	46.4	15.0	21.8
0.5027	0.4474	0.1	3.0	10	60.4	24.2	37.6
0.4513	0.3903	0.1	2.0	20	39.0	46.2	24.6
0.5500	0.5000	0.1	2.5	20	72.6	76.0	54.0
0.6312	0.5902	0.1	3.0	20	87.4	87.2	74.8
0.5109	0.4565	0.1	2.0	30	27.6	68.2	41.0
0.6108	0.5675	0.1	2.5	30	54.0	90.0	69.8
0.6886	0.6540	0.1	3.0	30	78.6	96.8	86.6
0.4853	0.2647	0.3	2.0	10	22.8	5.40	11.4
0.5519	0.3599	0.3	2.5	10	45.6	16.8	28.2
0.6132	0.4474	0.3	3.0	10	64.6	27.4	50.4
0.5732	0.3903	0.3	2.0	20	42.0	42.6	30.8
0.6500	0.5000	0.3	2.5	20	72.2	73.0	57.4
0.7131	0.5902	0.3	3.0	20	87.2	88.0	79.8
0.6196	0.4565	0.3	2.0	30	32.4	70.6	51.0
0.6973	0.5675	0.3	2.5	30	64.8	90.8	75.0
0.7578	0.6540	0.3	3.0	30	82.0	98.0	89.8
0.6324	0.2647	0.5	2.0	10	23.2	9.60	13.6
0.6800	0.3599	0.5	2.5	10	44.2	18.2	30.8
0.7237	0.4474	0.5	3.0	10	60.0	28.6	52.4
0.6952	0.3903	0.5	2.0	20	38.2	42.0	39.2
0.7500	0.5000	0.5	2.5	20	72.6	72.2	68.6
0.7951	0.5902	0.5	3.0	20	84.0	87.2	80.4
0.7283	0.4565	0.5	2.0	30	39.0	71.4	60.2
0.7838	0.5675	0.5	2.5	30	67.2	92.4	84.2
0.8270	0.6540	0.5	3.0	30	83.0	97.2	94.2

^aAnalysis model had a prevalence of 10

^bAnalysis model had a prevalence of 30

^cVariance components

et al. 2000) was used for VC analyses. After LINKAGE was used to estimate lod scores for a range of recombination fractions (0, 0.01, 0.05, 0.1, 0.2, 0.3, and 0.4), the maximum lod score across the different recombination fractions for each replicate was used to estimate the power. Power was defined as the percentage of replicates above a given lod score threshold. We estimated power for lod > 2 (suggestive linkage) and lod > 3 (significant linkage). Power estimates are presented only for the latter situation.

Results and discussion

Power estimates

Tables 1, 2, 3, and 4 summarize analyses of the different simulation models, displaying a wide range of estimated power to achieve a lod score greater than 3. Table columns summarize (1) parameter values assigned in the simulation model, (2) power estimates obtained for PSL analysis assuming prevalence of either 10% or 30%, and (3) power estimate obtained by VC analysis. The power reported for the parametric analyses was based on maximum lod scores across a range of recombination fractions for the simulated two-point and fully informative models.

The power estimates varied widely as a function of the underlying prevalence, displacement, and QTL variance (and therefore total heritability) assumed in the data-generating models. As anticipated, displacement is a critical determinant of the power for both PSL and VC models. For the two-point models, when the simulation and analysis models were both dominant with the same prevalence, power for lod > 3 ranged from 44% to 92% for a displacement of 2.5 with QTL variance ranging from 0.36 to 0.57 (total heritability 0.42–0.78). For a displacement of 3, power for lod > 3 ranged from 60% to 98% with a QTL variance range of 0.45–0.65 (total heritability 0.50–0.83) (Tables 1, 2). The power of multipoint linkage analysis was approximated using a tightly linked, fully informative marker (Tables 3, 4). As expected from formal analyses, full marker information led to a marked gain in power in all PSL and VC analyses. When the simulation and analysis models were both dominant with the same prevalence, power for lod > 3 ranged from 64% to 99% for a displacement of 2.5 with QTL variance ranging from 0.36 to 0.57 (total heritability 0.42–0.78). For a displacement of 3, power for lod > 3 ranged from 81.4% to 100% with a QTL variance range of 0.45–0.65 (total heritability 0.50–0.83).

Comparing data generated under a dominant or a recessive model for the same displacement and QTL

Table 2 Power estimates (%) for lod > 3. Simulation model: recombination fraction of $\theta=0.05$, two-point, recessive; analysis model: recessive. *QTL* quantitative trait locus

Simulation						Analysis		
Total heritability	QTL variance	Residual heritability	<i>t</i> (displacement)	Prevalence		Power 10 ^a	Power 30 ^b	Power VC ^c
0.3382	0.2647	0.1	2.0	10		14.8	5.40	4.60
0.4239	0.3599	0.1	2.5	10		38.2	10.6	12.4
0.5027	0.4474	0.1	3.0	10		56.4	21.0	28.2
0.4513	0.3903	0.1	2.0	20		32.4	34.4	18.6
0.5500	0.5000	0.1	2.5	20		64.4	66.4	44.0
0.6312	0.5902	0.1	3.0	20		82.2	87.0	65.4
0.5109	0.4565	0.1	2.0	30		20.4	63.6	36.8
0.6108	0.5675	0.1	2.5	30		50.4	90.8	66.8
0.6886	0.6540	0.1	3.0	30		74.2	96.8	86.2
0.4853	0.2647	0.3	2.0	10		14.8	3.60	5.60
0.5519	0.3599	0.3	2.5	10		39.2	12.2	19.6
0.6132	0.4474	0.3	3.0	10		58.0	23.8	39.8
0.5732	0.3903	0.3	2.0	20		34.2	37.6	28.2
0.6500	0.5000	0.3	2.5	20		65.2	67.4	55.8
0.7131	0.5902	0.3	3.0	20		82.0	85.2	73.4
0.6196	0.4565	0.3	2.0	30		25.4	60.8	42.8
0.6973	0.5675	0.3	2.5	30		58.2	91.4	78.6
0.7578	0.6540	0.3	3.0	30		73.0	97.2	86.0
0.6324	0.2647	0.5	2.0	10		14.6	6.80	11.0
0.6800	0.3599	0.5	2.5	10		34.4	14.6	24.8
0.7237	0.4474	0.5	3.0	10		54.8	25.2	45.4
0.6952	0.3903	0.5	2.0	20		35.4	38.0	32.8
0.7500	0.5000	0.5	2.5	20		65.6	69.2	61.0
0.7951	0.5902	0.5	3.0	20		81.8	82.4	78.8
0.7283	0.4565	0.5	2.0	30		32.0	65.8	55.4
0.7838	0.5675	0.5	2.5	30		58.6	88.6	77.8
0.8270	0.6540	0.5	3.0	30		76.0	97.8	91.2

^aAnalysis model had a prevalence of 10

^bAnalysis model had a prevalence of 30

^cVariance components

variance, the lower the population prevalence, the lower was the power when a recessive model was assumed for both the simulation and linkage analysis. This reflects the fact that, at lower population prevalences, the proportion of informative families and the proportion of segregating offspring is greater for a dominant than for a recessive trait. As prevalence increases, the behavior of the two models becomes more similar due to the manner in which these two proportions are differentially affected by these two modes of inheritance. Misspecifying the dominance of the trait is known to markedly reduce the statistical power of parametric methods of analysis (Haines and Pericak-Vance 1998). In the present instance, we found that the power was less than 20% in all cases where this relationship was misspecified in the analysis (data not shown).

Parametric models not only require an assumption about dominance but also depend on assumptions about the prevalence. Our analyses consistently showed that the loss of power was more pronounced when prevalence was overestimated than when it was underestimated in the analysis (Tables 1, 2, 3, 4). A similar observation was made by Pal et al. (2001).

The relative power of PSL and VC models could also be evaluated from Tables 1, 2, 3 and 4. As expected, PSL models had markedly greater power than VC models when the parameters used in the analysis were close to

those used to simulate the data. The robustness of VC models, however, was reflected in the greater power they achieved compared to PSL models when the dominance relationship was misspecified. A note of caution is in order regarding these conclusions, however, since the distribution of type I and type II errors may not be similar for the two approaches.

Another feature revealed by inspection of Tables 1, 2, 3 and 4 was that residual heritability, while it had little impact on the power of PSL models, led to a proportional increase in power of VC methods, particularly when a dominant model was assumed in the simulation models.

Extrapolation of power for the full set of CEPH families

Whereas simulations were performed for the 30 pedigrees sampled in the initial phase of the reinterview process, 45 families are expected to participate in the UGRP. To estimate the gain in power that could be expected from this sample size increase, the empirical distribution of lod scores for each set of replicates was examined. For PSL models of analysis, the square root of the lod scores followed approximately a normal distribution. We therefore used a normal approximation to assess the gain in power; the VC lod scores did not

Table 3 Power estimates (%) for lod > 3: Simulation model: recombination fraction of $\theta=0.005$, fully informative, dominant; analysis model: dominant. *QTL* quantitative trait locus

Simulation						Analysis		
Total heritability	QTL variance	Residual heritability	<i>t</i> (displacement)	Prevalence		Power 10 ^a	Power 30 ^b	Power VC ^c
0.3382	0.2647	0.1	2.0	10		38.8	11.4	10.6
0.4239	0.3599	0.1	2.5	10		73.0	27.8	35.2
0.5027	0.4474	0.1	3.0	10		84.2	42.8	57.8
0.4513	0.3903	0.1	2.0	20		58.6	64.6	43.8
0.5500	0.5000	0.1	2.5	20		85.6	87.2	70.4
0.6312	0.5902	0.1	3.0	20		98.0	98.6	89.6
0.5109	0.4565	0.1	2.0	30		44.6	89.6	63.8
0.6108	0.5675	0.1	2.5	30		78.4	99.2	90.0
0.6886	0.6540	0.1	3.0	30		93.4	99.6	97.4
0.4853	0.2647	0.3	2.0	10		38.0	13.4	19.4
0.5519	0.3599	0.3	2.5	10		69.8	27.8	39.6
0.6132	0.4474	0.3	3.0	10		81.4	40.6	61.2
0.5732	0.3903	0.3	2.0	20		62.0	63.6	51.2
0.6500	0.5000	0.3	2.5	20		88.8	88.4	80.4
0.7131	0.5902	0.3	3.0	20		97.4	97.4	92.4
0.6196	0.4565	0.3	2.0	30		50.4	87.2	70.6
0.6973	0.5675	0.3	2.5	30		84.2	98.6	92.2
0.7578	0.6540	0.3	3.0	30		94.8	100	98.8
0.6324	0.2647	0.5	2.0	10		36.2	14.6	20.4
0.6800	0.3599	0.5	2.5	10		64.0	28.8	47.4
0.7237	0.4474	0.5	3.0	10		82.8	44.0	72.8
0.6952	0.3903	0.5	2.0	20		63.2	65.4	60.0
0.7500	0.5000	0.5	2.5	20		90.4	89.0	88.6
0.7951	0.5902	0.5	3.0	20		97.6	97.2	96.6
0.7283	0.4565	0.5	2.0	30		59.6	90.0	80.6
0.7838	0.5675	0.5	2.5	30		85.6	99.0	97.4
0.8270	0.6540	0.5	3.0	30		95.4	99.8	98.8

^aAnalysis model had a prevalence of 10

^bAnalysis model had a prevalence of 30

^cVariance components

Table 4 Power estimates (%) for lod > 3. Simulation model: recombination fraction of $\theta = 0.005$, fully informative, recessive; analysis model: recessive. *QTL* quantitative trait locus

Simulation					Analysis		
Total heritability	QTL variance	Residual heritability	t (displacement)	Prevalence	Power 10 ^a	Power 30 ^b	Power VC ^c
0.3382	0.2647	0.1	2.0	10	32.6	7.40	6.40
0.4239	0.3599	0.1	2.5	10	61.4	18.8	26.0
0.5027	0.4474	0.1	3.0	10	81.4	32.2	45.6
0.4513	0.3903	0.1	2.0	20	50.2	57.0	32.8
0.5500	0.5000	0.1	2.5	20	87.0	88.4	67.2
0.6312	0.5902	0.1	3.0	20	95.4	97.6	86.2
0.5109	0.4565	0.1	2.0	30	39.8	84.6	59.6
0.6108	0.5675	0.1	2.5	30	76.2	98.8	86.6
0.6886	0.6540	0.1	3.0	30	91.8	99.6	96.4
0.4853	0.2647	0.3	2.0	10	29.0	8.00	11.8
0.5519	0.3599	0.3	2.5	10	60.8	19.2	31.4
0.6132	0.4474	0.3	3.0	10	79.8	38.8	54.4
0.5732	0.3903	0.3	2.0	20	55.6	57.6	45.8
0.6500	0.5000	0.3	2.5	20	85.0	87.6	73.2
0.7131	0.5902	0.3	3.0	20	95.8	95.8	89.4
0.6196	0.4565	0.3	2.0	30	47.2	85.2	68.6
0.6973	0.5675	0.3	2.5	30	78.2	98.6	91.4
0.7578	0.6540	0.3	3.0	30	93.2	100	99.2
0.6324	0.2647	0.5	2.0	10	26.0	9.40	20.6
0.6800	0.3599	0.5	2.5	10	59.2	25.6	42.0
0.7237	0.4474	0.5	3.0	10	74.2	38.6	61.6
0.6952	0.3903	0.5	2.0	20	55.8	56.4	51.0
0.7500	0.5000	0.5	2.5	20	87.8	88.4	84.8
0.7951	0.5902	0.5	3.0	20	95.4	94.4	92.6
0.7283	0.4565	0.5	2.0	30	55.4	86.0	76.4
0.7838	0.5675	0.5	2.5	30	82.4	98.2	95.4
0.8270	0.6540	0.5	3.0	30	95.2	100	99.0

^aAnalysis model had a prevalence of 10^bAnalysis model had a prevalence of 30^cVariance components

distribute normally and were not included in the approximation. Since all the Utah CEPH families have similar pedigree structures, it was reasonable to assume that the families in the current sample of 30 pedigrees are representative of the 15 more to be added.

The mean and variance of the lod score distribution for a given analysis model were multiplied by 1.5 (since the sample size will increase 1.5 times) using the principles of statistics that state $E(cX) = cE(X)$ and $\text{Var}(cX) = c^2\text{Var}(X)$, where c is a constant, $E(X)$ is the expected value (mean) of X , and $\text{Var}(X)$ is the variance of X . Therefore, in our analysis, the mean was multiplied by 1.5 and the variance was multiplied by $(1.5)^2$. Using these 45-pedigree means (μ) and variances (σ^2), a standard normal equation $(\sqrt{\text{lod}} - \mu)/\sigma$ (where $\sqrt{\text{lod}}$ = $\sqrt{\text{cut-off lod score}}$ which, in our case, equals 3) was then referred to a table of the normal density function. The results of these power calculations for an extended sample of 45 pedigrees are presented in Table 5. The average power increase varied from 17% to 26% with maximum values ranging from 34% to 40%.

Conclusion

The analyses presented here estimate the power that can be achieved for linkage studies performed in the Utah

CEPH pedigrees as a function of the parameters that characterize the distribution of a quantitative trait. The tables show parameter values likely to provide adequate power prior to analysis whether PSL or VC methods are to be used. Although a QTL variance < 26%, total heritability < 35%, and a displacement < 2 gave low power estimates for lod > 3, values above these limits yielded adequate power (a large number of models giving power > 85%). Trait parameters that play a large role in detecting linkage are prevalence, displacement, QTL variance, and total heritability of the trait. As the focus shifts to intermediate phenotypes in the genetic analysis of common disease, this analysis documents that the use of a reference set with large sibship size such

Table 5 Average and maximum increase (%) in power (lod > 3) with the addition of 15 pedigrees to the sample. Results are based on parametric analysis models

	Mean increase	Maximum increase
Two-point model		
Dominant	17.92	35.34
Recessive	25.68	39.57
Fully informative model		
Dominant	16.95	34.44
Recessive	18.98	39.01

as the Utah CEPH families can provide considerable statistical power to detect linkage for common traits.

Acknowledgements This investigation was supported by a Public Health Services research grant to the Huntsman General Clinical Research Center at the University of Utah, Research Grant M01-RR00064 from the National Center for Research Resources, and NIH grant HL067469. It was also supported by generous gifts from the W.M. Keck Foundation and from the George S. and Delores Doré Eccles Foundation. We would like to extend our sincere thanks to all family members who participated in the Utah Genetic Reference Project. Thanks also to Andreas P. Peiffer, UGRP Medical Director, and Melissa M. Dixon, UGRP Study Coordinator. Jean-Marc Lalouel provided valuable comments and suggestions during the preparation of this manuscript.

References

- Blangero J, Williams JT, Almasy L (2000) Quantitative trait locus mapping using human pedigrees. *Hum Biol* 72:35–62
- Boehnke M (1986) Estimating the power of a proposed linkage study: a practical computer simulation approach. *Am J Hum Genet* 39:513–527
- Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R (1990) Centre d'étude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 6:575–577
- Duggirala R, Williams JT, Williams-Blangero S, Blangero J (1997) A variance component approach to dichotomous trait linkage analysis using a threshold model. *Genet Epidemiol* 14:987–992
- Ekstrom CT (2001) Power of multipoint identity-by-descent methods to detect linkage using variance component models. *Genet Epidemiol* 21:285–298
- Elston RC (1979) Major locus analysis for quantitative traits. *Am J Hum Genet* 31:655–661
- Greenberg DA, Abreu P, Hodge SE (1998) The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am J Hum Genet* 63:870–879
- Haines JL, Pericak-Vance MAE (1998) Approaches to gene mapping in complex human diseases. Wiley-Liss, New York
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3–19
- Hasstedt S (2002) PAP: Pedigree Analysis Package, v. 5. Department of Human Genetics, University of Utah, Salt Lake City
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347–1363
- Lathrop GM, Lalouel JM, Julier C, Ott J (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *Am J Hum Genet* 37:482–498
- Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland
- Ott J (1999) Analysis of human genetic linkage. The Johns Hopkins University Press, Baltimore
- Pal DK, Durner M, Greenberg DA (2001) Effect of misspecification of gene frequency on the two-point LOD score. *Eur J Hum Genet* 9:855–859
- Ploughman LM, Boehnke M (1989) Estimating the power of a proposed linkage study for a complex genetic trait. *Am J Hum Genet* 44:543–551
- Pratt SC, Daly MJ, Kruglyak L (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 66:1153–1157
- Rijsdijk FV, Hewitt JK, Sham PC (2001) Analytic power calculation for QTL linkage analysis of small pedigrees. *Eur J Hum Genet* 9:335–340
- Risch N (1997) Evolving methods in genetic epidemiology. II. Genetic linkage from an epidemiologic perspective. *Epidemiol Rev* 19:24–32
- Sham P (1998) Statistics in human genetics. Arnold, London
- Shugart YY, O'Connell JR, Wilson AF (2002) An evaluation of the variance components approach: type I error, power and size of the estimated effect. *Eur J Hum Genet* 10:133–136
- The Utah Marker Development Group (1995) A collection of ordered tetranucleotide-repeat markers from the human genome. *Am J Hum Genet* 57:619–628
- White R, Leppert M, Bishop DT, Barker D, Berkowitz J, Brown C, Callahan P, Holm T, Jerominski L (1985) Construction of linkage maps with DNA markers for human chromosomes. *Nature* 313:101–105
- Wijmsman EM, Amos CI (1997) Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. *Genet Epidemiol* 14:719–735
- Williams JT, Blangero J (1999) Power of variance component linkage analysis to detect quantitative trait loci. *Ann Hum Genet* 63:545–563