Statistical Power of Expression Quantitative Trait Loci for Mapping of Complex Trait Loci in Natural Populations

Paul Schliekelman¹

Department of Statistics, University of Georgia, Athens, Georgia 30602

Manuscript received May 29, 2007

Accepted for publication January 11, 2008

ABSTRACT

A number of recent genomewide surveys have found numerous QTL for gene expression, often with intermediate to high heritability values. As a result, there is currently a great deal of interest in genetical genomics—that is, the combination of genomewide expression data and molecular marker data to elucidate the genetics of complex traits. To date, most genetical genomics studies have focused on generating candidate genes for previously known trait loci or have otherwise leveraged existing knowledge about trait-related genes. The purpose of this study is to explore the potential for genetical genomics approaches in the context of genomewide scans for complex trait loci. I explore the expected strength of association between expression-level traits and a clinical trait, as a function of the underlying genetic model in natural populations. I give calculations of statistical power for detecting differential expression between affected and unaffected individuals. I model both reactive and causative expression-level traits with both additive and multiplicative multilocus models for the relationship between phenotype and genotype and explore a variety of assumptions about dominance, number of segregating loci, and other parameters. There are two key results. If a transcript is causative for the disease (in the sense that disease risk depends directly on transcript level), then the power to detect association between transcript and disease is quite good. Sample sizes on the order of 100 are sufficient for 80% power. On the other hand, if the transcript is reactive to a disease locus, then the correlation between expression-level traits and disease is low unless the expression-level trait shares several causative loci with the disease—that is, the expression-level trait itself is a complex trait. Thus, there is a trade-off between the power to show association between a reactive expression-level trait and the clinical trait of interest and the power to map expression-level QTL (eQTL) for that expression-level trait. Gene expression-level traits that are most strongly correlated with the clinical trait will themselves be complex traits and therefore often hard to map. Likewise, the expression-level traits that are easiest to map will tend to have a low correlation with the clinical trait. These results show some fundamental principles for understanding power in eQTL-based mapping studies.

IT has become apparent that mapping of complex trait loci is a much more difficult task than once envisioned. GLAZIER *et al.* (2002) reviewed the literature for major experimental species and found only a relative handful of complex traits whose genetic basis is well understood. ALTMULLER *et al.* (2001) conducted a comprehensive review of 101 whole-genome scan studies of complex diseases in humans and found that only one-third produced significant linkages. Furthermore, few of the linkages that were significant have been reproduced in other studies. In contrast, geneticists have been very successful at mapping single-locus traits. Hundreds of single-locus traits have been mapped in humans alone.

The reason that single-locus traits are relatively easy to map is that there is a strong correlation between genotype at the causative locus and the trait. If a qualitative trait is present, then there is usually or always a

corresponding genotype at the causative locus. By contrast, with a multilocus trait it is possible to exhibit the trait without there being a “trait genotype” at a particular locus or to not exhibit the trait when there is a trait genotype at that causative locus. Thus, there is not a consistent genotypic difference between those individuals exhibiting the trait and those who are not (or between different levels of the trait for a quantitative trait) and power to map complex trait loci is low.

In the past few years, the new field of “genetical genomics” has emerged—that is, the combination of molecular marker data and genomewide expression data to elucidate the genetic basis of gene expression. Evidence has emerged of substantial genetic variation in gene expression (*e.g.*, BREM *et al.* 2002; OLEKSIK *et al.* 2002; YAN *et al.* 2002; CHEUNG *et al.* 2003; SCHADT *et al.* 2003; MORLEY *et al.* 2004). For example, BREM *et al.* (2002) conducted linkage analysis of expression levels in a cross between two yeast strains. They found 1528 genes with differential expression. Among those genes they found a median heritability of expression of 84%.

¹Author e-mail: pdschlie@stat.uga.edu

Expression levels for 308 of these genes showed linkage to at least one marker. MONKS *et al.* (2004) measured expression in 23,499 genes in lymphoblastoid cell lines from members of 15 CEPH families. A total of 2340 were found to be differentially expressed and 762 (31%) of these had heritability significantly different than zero. Median heritability was 0.34 for these genes and 25% had heritability >0.44 . A genome scan was performed and 55 expression-level QTL (eQTL) for 33 genes were found. Twenty of these QTL explained $>90\%$ of the variance and >45 explained $>70\%$ of the variance.

These and other studies show that there may be many expression-level traits with a relatively simple and tractable genetic basis. This indicates that microarrays might be a powerful tool for mapping loci underlying complex traits. The premise is that certain gene expression-level traits might have a stronger correlation with the underlying causative loci than the trait does itself. That is, (1) many expression-level traits are highly heritable relative to complex clinical traits and appear to have a relatively simple genetic basis, (2) there is evidence of multiple expression-level traits mapping to the same chromosomal region (and perhaps the same locus), and (3) if genetic polymorphisms that affect disease risk also affect expression levels of other genes, these expression-level traits will give high power for mapping the polymorphisms. By causative locus, I mean a locus at which there is genotypic variation responsible for variation in a clinical trait of interest. By clinical trait, I mean an organismal trait of interest, such as disease status, body weight, etc. Of course, the word "clinical" is used just for convenience—the trait need not solely be of clinical interest.

The potential of this type of approach was demonstrated by SCHADT *et al.* (2003). They performed an F_2 cross between two inbred mouse strains and performed a genomewide scan for linkage for expression levels of 23,574 genes. The mice were on a high-fat diet for 4 months and thus developed a spectrum of obesity. They were classified according to the subcutaneous fat-pad-mass (FPM) trait and expression levels were compared between the upper 25% and the lower 25% for this trait. They examined the 280 genes most differentially expressed between these two groups and found that clustering on this set of genes divided the mice into two high-FPM groups and one low-FPM group, revealing genetic heterogeneity in the high-FPM mice. They then performed a genome scan for the FPM trait and showed that LOD scores of QTL for this trait were substantially increased when only one of the high-FPM groups was included. They also showed that a number of genes had expression-level traits mapping to the same region as a QTL for the FPM trait.

To apply genetical genomics approaches to a clinical trait (*e.g.*, a disease), it is necessary to establish a connection between expression-level traits and the clinical trait. To date, most of the focus in genetical genomics

has been in exploring the quantitative genetics of gene expression and in candidate gene studies. Substantial progress has been made, on both experimental (*e.g.*, the articles cited above) and statistical (see KENDZIORSKI and WANG 2006 for review) fronts. However, little is still known about how transcriptional variation among genotypes relates to variation in clinical traits (GIBSON and WEIR 2005). Until this is better understood, it will be difficult to apply genetical genomics approaches to understanding complex traits. The goal of this article is to explore how transcriptional variation among genotypes should be expected to relate to variation in a qualitative trait.

The ideal expression-level trait for mapping would be one whose variation is entirely determined by the genotype at a single one of the L causative loci. If such expression-level traits existed for each of the causative loci, then they could all be mapped with relative ease. However, such expression-level traits would not correlate strongly with the disease and might be difficult to detect as differentially expressed between disease affected and unaffected individuals. Expression-level traits that correlate strongly with the disease would be easiest to detect as differentially expressed, but would not be any better than the disease itself as a response variable for mapping. SCHADT *et al.* (2003) and others have found a very large number of eQTL mapping to locations throughout the genome, irrespective of trait status. Determining which of these eQTL are relevant requires establishing a correlation between the expression-level trait and the disease. Thus, the use of eQTL for mapping complex traits may just shift the obstacle of low power for establishing a correlation between trait and genotype to a new obstacle of low power for establishing a correlation between disease and expression level. Thus, it is unclear how powerful we should expect eQTL-based strategies to be.

A number of studies have used microarrays for identifying candidate genes (*e.g.*, LAWN *et al.* 1999; BERGE *et al.* 2000; KARP *et al.* 2000; WAYNE and MCINTYRE 2002; KIRST *et al.* 2004; BYSTRYKH *et al.* 2005; CHESLER *et al.* 2005; HUBNER *et al.* 2005; LI *et al.* 2005; SCHADT *et al.* 2005). For example, BYSTRYKH *et al.* (2005) mapped eQTL from mouse stem cells. Those that mapped to the region of a known QTL for stem cell turnover were then identified as candidate genes. Such approaches are promising when there is prior knowledge of the location of a causative locus. The primary focus of this article is genome scans—that is, whole-genome linkage or association studies when there is no candidate region previously identified.

There are two basic tasks that must be accomplished to use genomewide expression information for a genome scan: (1) an association must be established between the transcript and the clinical trait and (2) association must be established between the transcript and a marker. These tasks might be done in a two-stage analysis or as some combined analysis, but they must be accomplished somehow. First, I examine the statistical

power to detect expression-level changes as being associated with the clinical trait. Then, I examine the power to detect eQTL for expression-level traits that have been identified as differentially expressed.

I consider both transcripts that are disease causative (that is, variation in transcript level contributes directly to disease risk) and those that are controlled by a disease locus but are not causative. These latter transcripts are termed “reactive.” Note that this term refers to transcripts reactive to a disease locus and not to the disease itself (thus the usage is somewhat different from that in some previous eQTL articles). Power is shown to be quite good for causative transcripts, with sample sizes on the order of 100 being sufficient for reliable detection of association with the clinical trait. The story is more complicated for reactive transcripts. There is a trade-off in power between the two tasks laid out in the previous paragraph. The sample sizes needed for reasonable power to detect expression-level differences are often very high unless the expression-level trait and disease share multiple causative loci (thus creating a higher correlation between disease and expression level). Unfortunately, the power to map these causative loci via the expression-level trait plummets as the number of influencing loci increases. The result of this trade-off is that, for natural populations, there may be no expression-level traits that will give good power for detecting association both with the clinical trait and with a marker (although the exact power varies greatly depending on model assumptions and parameter values). It should be stressed that I am considering only associations involving single expression-level traits and there is potentially a great deal of information contained in the joint behavior of expression-level traits. Furthermore, improved experimental designs are possible. Thus, it is not the goal of this article to make absolute statements about statistical power. Rather, the goal is to determine the fundamental issues and to provide a baseline for understanding power in such studies.

POWER TO DETECT EXPRESSION-LEVEL DIFFERENCES DUE TO POLYMORPHISMS BETWEEN AFFECTED AND UNAFFECTED INDIVIDUALS: METHODS

Trait and expression-level models: The focus of this article is on qualitative traits. I assume that the trait is binary and is controlled by genotype at L causative loci. For ease of description I refer to the trait as a disease. However, the results are equally valid for any binary trait. I further assume that a total of N expression-level traits $X_1, X_2, X_3, \dots, X_N$ are being tested. Each expression-level trait (ELT) depends on some subset of the L causative loci, with the subset being empty for the many expression-level traits that are independent of all of the causative loci. ELTs that are controlled by disease loci are assumed to have distributions that are mixtures of normal dis-

tributions. That is, the expression distribution is assumed to be normal distribution on some scale (e.g., log scale) with mean and variance dependent on the genotype at the controlling loci. It is presently uncertain how well this assumption holds for most expression levels, but this seems reasonable as a first approximation. Hsieh *et al.* (2007) found evidence of multimodality in transcription in 7–10% of genes in two *Drosophila melanogaster* lines. Other genes may have had either little within-population variation or genetic control too complicated to be detected with the available sample sizes.

Consider an ELT X_1 that is either causative for the disease or reactive to a disease locus. We wish to find the distribution for X_1 as a function of an individual's disease-affected status. If T is an indicator variable for disease status ($T = 1$ indicating disease and $T = 0$ indicating no disease), then we have

$$\begin{aligned} f(X_1 | T = 1) &= \frac{f(X_1, T = 1)}{P(T = 1)} \\ &= \frac{\sum_{\vec{g}} P(T = 1 | X_1, \vec{g}) f(X_1, \vec{g})}{K} \\ &= \frac{\sum_{\vec{g}} P(T = 1 | X_1, \vec{g}) f(X_1 | \vec{g}) P(\vec{g})}{K}, \end{aligned} \quad (1)$$

where $\vec{g}$ is a vector of genotypes at the causative loci and the disease prevalence is $K = P(T = 1)$.

Similarly,

$$\begin{aligned} f(X_1 | T = 0) &= \frac{\sum_{\vec{g}} (1 - P(T = 1 | X_1, \vec{g})) f(X_1 | \vec{g}) P(\vec{g})}{1 - K}. \end{aligned} \quad (2)$$

We can see that expression-level distributions will be mixtures of normal distributions, with the weights determined by the trait value and the population genetic parameters.

Multiplicative trait model: To proceed, we need to further specify the trait penetrance model. The first example model is a multiplicative one. If the ELT is reactive only, we have

$$P(T = 1 | X_1, \vec{g}) = u_1(g_1) u_2(g_2) \dots u_L(g_L), \quad (3)$$

where $u_j(g_j)$ is the contribution to disease risk from the one-locus genotype i_j on locus j . Thus, G is a vector $[g_1, g_2, \dots, g_L]$. The mechanistic basis for a multiplicative model is that each locus affects one of a chain of L disease events that must all occur for the disease to occur. By “disease event” I mean some physiological event associated with the locus that contributes to the occurrence of the disease.

Substituting (3) for $P(T = 1 | G)$ in (1),

$$\begin{aligned} f(X_1 | T = 1) &= \frac{\sum_{g_1, g_2, \dots, g_L} u_1(g_1) u_2(g_2) \dots u_L(g_L) f(X_1 | g_1, g_2, \dots, g_L) P(g_1) P(g_2) \dots P(g_L)}{K}, \end{aligned} \quad (4)$$

where $\sum_{g_1, g_2, \dots, g_L}$ denotes the multiple summations over the individual-locus genotypes g_i and I have suppressed the $X_1 = x$ and $G = g$ notation to simplify the appearance of the equations. I assume no gametic disequilibrium so that $P(\vec{g}) = P(g_1)P(g_2) \dots P(g_L)$. In this case $K = K_1 K_2 \dots K_L$, where

$$K_i = \sum_{g_i} P(g_i) u_i(g_i). \quad (5)$$

Thus (4) can be simplified to

$$f(X_1 | T = 1) = \frac{\sum_{g_1, g_2, \dots, g_c} f(X_1 | g_1, g_2, \dots, g_c) u_1(g_1) u_2(g_2) \dots u_c(g_c) P(g_1) P(g_2) \dots P(g_c)}{K_1 K_2 \dots K_c}, \quad (6)$$

where I assume, without loss of generality, that the loci have been indexed such that the loci 1– c are the ones that affect the expression level X_1 . Thus, the expression-level distribution for affected individuals depends only on genotypes at the loci influencing that expression-level trait. On the other hand, the factor $1 - K$ in the denominator of (2) cannot be separated by genotype and thus no cancellation occurs and the expression-level distribution is dependent on genotype at all L causative loci. Thus, the distribution (1) will have a strong dependence on the genotype at the loci 1, $\dots$, c , while this dependence will be obscured by effects from the other loci in the distribution (2). This reflects the fact that under a multiplicative model an affected individual has always experienced a disease event at all L causative loci while an unaffected individual may have disease alleles at particular loci (he/she just does not have disease genotypes at all disease loci).

If the ELT is causative, the disease risk depends both on genotype and on the value of the ELT X_1 ,

$$P(T = 1 | X_1, \vec{g}) = u(X_1) u_{c+1}(g_{c+1}) \dots u_L(g_L), \quad (7)$$

where $u(X_1)$ is the contribution to disease risk from the ELT X_1 . I assume that loci 1– c affect disease risk only through their impact on X_1 . Equation 2 becomes

$$f(X_1 | T = 1) = \frac{\sum_{g_1, g_2, \dots, g_c} u(X_1) f(X_1 | g_1, g_2, \dots, g_c) P(g_1) P(g_2) \dots P(g_c)}{K / (K_{c+1} \dots K_L)}. \quad (8)$$

The functions $u_{c+1}(g_{c+1}) \dots u_L(g_L)$ are of the same type as those for the reactive model. Note that it is not specified whether there are causative ELTs associated with the loci $c + 1$ – L . The effects of these loci are contained in the functions $u_j(g_j)$. The effect could be via a causative ELT or via some other mechanism.

Additive trait model: Under an additive model with a reactive ELT, the disease probability is given by

$$u(\vec{g}) = u_1(g_1) + u_2(g_2) \dots + u_L(g_L). \quad (9)$$

Under this model the disease event associated with any causative locus is sufficient to cause the disease, but the more that such disease events occur the higher the chance of disease. Equation 1 becomes

$$f(X_1 | T = 1) = \frac{\sum_{g_1} u_1(g_1) P(g_1) f(X_1 | g_1) + \dots + \sum_{g_c} u_c(g_c) P(g_c) f(X_1 | g_c) + f(X_1)(K_{c+1} + \dots + K_L)}{K} \quad (10)$$

and (2) becomes

$$f(X_1 | T = 0) = \frac{f(X_1) - \sum_{g_1} u_1(g_1) P(g_1) f(X_1 | g_1) - \dots - \sum_{g_c} u_c(g_c) P(g_c) f(X_1 | g_c) - f(X_1)(K_{c+1} + \dots + K_L)}{1 - K}, \quad (11)$$

where $K = K_1 + K_2 + \dots K_L$ under an additive model. Similar expressions apply for a causative ELT.

Expression-level distribution: I use the above relationships to calculate the power to detect expression-level differences due to genotypic differences between affected and unaffected individuals. The premise is that there should be a large number of changes in expression-level traits associated with the occurrence of a disease or other trait.

Suppose that some individual has mutations at some set of disease causative loci and that each such mutation initiates some sequence of events. These different sequences of events will then eventually interact in some way to cause the disease (or, more generally, to increase disease probability). ELTs that are “upstream” in the sequence of events originating from a mutation will have a strong association with that mutation, but a weak association with the disease. That is, if the mutation is weakly associated with the disease (as is the case for complex diseases), then ELTs strongly associated with that mutation must also be weakly associated with the disease. On the other hand, ELTs that occur “downstream” of the interactions between mutations will have stronger association with the disease but weaker association with the disease loci (because these ELTs are themselves complex traits). I assume that no reactive ELTs are simultaneously upstream and downstream—that is, for example, that no reactive ELT depends directly both on genotype at a causative locus and on disease status (although there would be indirect dependencies in both cases). The precise meaning of this becomes clear in the derivations that follow.

My focus is on the statistical power to map a specific one of the L causative loci. I call this locus the target causative locus (TCL). I assume that the genotype at this causative locus affects the expression level of M genes (possibly including itself). If there is a causative ELT, then M is always taken as one. The genotype at loci besides the L causative loci may also affect an expression-level trait, but such loci are uncorrelated with the disease and their effect is assumed to be included in the

normal distributions. When I refer to a causative locus as influencing or controlling an expression level, I mean specifically that genetic variation at the causative locus affects the expression-level trait.

Under the assumptions stated earlier, the expression-level distribution is a mixture of normal distributions, where the mean and the variance of those normals are genotype dependent. For $c = 1$, and a multiplicative model with a reactive ELT, we have

$$f(X_1 | T = 1) = \frac{\sum_{g_1} N(x; \mu_{g_1}, \sigma_{g_1}) u_1(g_1) P(g_1)}{K_1} \quad (12)$$

and

$$\begin{aligned} f(X_1 | T = 0) &= \frac{\sum_{g_1} N(x; \mu_{g_1}, \sigma_{g_1}) (1 - u_1(g_1) K_2 K_3 \dots K_L) P(g_1)}{1 - K}, \end{aligned} \quad (13)$$

where $N(x; \mu_{g_1}, \sigma_{g_1})$ denotes a normal probability density distribution with mean μ_{g_1} and variance $\sigma_{g_1}^2$.

Standard statistical theory states that if the probability density function is

$$f(x) = p_1 N(x; \mu_1, \sigma_1) + p_2 N(x; \mu_2, \sigma_2) + \dots + p_L N(x; \mu_n, \sigma_n),$$

where the p_i sum to 1, then the mean and variance are given by

$$E(x) = \sum_{i=1}^L p_i \mu_i$$

and

$$\text{Var}(x) = \sum_{i=1}^L p_i \sigma_i^2 + \sum_{i=1}^L p_i (\mu_i - E(x))^2. \quad (14)$$

The p_i 's in this case refer to the conditional genotype probabilities like those occurring in Equations 12 and 13: that is, $u_1(g_1)P(g_1)/K_1$ and $(1 - u_1(g_1)K_2K_3 \dots K_L)P(g_1)/(1 - K)$. The resulting distribution is not normal. However, I consider sample sizes large enough that the central limit theorem applies and the distribution of means is normal. Power is then calculated using standard theory for the comparison of two groups.

If there are multiple expression-level traits dependent on the TCL (*i.e.*, $M > 1$), then X_1 becomes a vector of expression levels and the normal distribution is replaced by a multivariate normal. In this case a simulation approach is used for power calculations (described later).

We next need to specify how the expression mean and variance depend on genotype. I take the variance as a constant for all genotypes. I use either a multiplicative or an additive model for the dependence of the mean

on genotype. If expression is multiplicative and controlled by loci 1– c , then the expression mean is given by

$$\mu_g = \mu_{g_1} \mu_{g_2} \dots \mu_{g_c}. \quad (15)$$

The value μ_{g_i} is the contribution to the mean from locus i .

In the additive case, the expression mean is given by

$$\mu_g = \mu_{g_1} + \mu_{g_2} + \dots + \mu_{g_c}. \quad (16)$$

Except where noted otherwise, the form of genetic control (multiplicative or additive) is the same for expression level and disease.

Population genetic parameters: I assume that all alleles at the TCL can be classified either as normal or as disease causing. All disease-causing alleles at the TCL are assumed to have equal effect. The frequency of disease alleles (denoted by D) is p and that of normal alleles (denoted by d) is $1 - p$. The contribution to disease risk $u(g_i)$ is π_L for a normal homozygote (genotype dd), δ_L for a disease homozygote (genotype DD), and $\pi_L + h(\delta_L - \pi_L)$ for heterozygotes (genotype Dd). These are the probabilities that the disease event associated with that locus occurs, given genotype. The amount of disease risk associated with the TCL is given by a parameter L . The single-locus contribution to disease risk K_L is taken as equal to $K^{1/L}$ for a multiplicative model and K/L for an additive one. Thus, this is the contribution that locus would have if there were L loci all of equal effect. Note that this does not assume that there are actually L loci of equal effect, just that the effect attributable to the TCL is the same as if there were. Note that the role of L in the previous derivations is no different when we interpret L this way. For the target causative locus I take $\pi_L = \pi^{1/L}$ and $\delta_L = \delta^{1/L}$ for a multiplicative penetrance model and $\pi_L = \pi/L$ and $\delta_L = \delta/L$ for an additive model.

Very little is known about allele frequencies or other parameters for complex traits. Because the prevalence K is typically known, I take it as given and solve for the allele frequency p using (5) after specifying the parameters h , π , δ , and L .

The parameters μ_{2c} and μ_0 are the mean expression levels for individuals with two disease alleles at all c loci controlling the expression level and the genotype with 0 disease alleles at those loci, respectively. If the expression genetics are multiplicative, then the single-locus expression mean μ_{g_i} is given by $\mu_0^{1/L}$, $\mu_0^{1/L} + h_e(\mu_{2c}^{1/L} - \mu_0^{1/L})$, and $\mu_{2c}^{1/L}$ for individuals with zero, one, and two disease alleles at locus i , respectively. If expression is controlled additively, then these values become μ_0/L , $\mu_0/L + h_e(\mu_{2c}/L - \mu_0/L)$, and μ_{2c}/L .

The parameter σ_E is the standard deviation in expression (assumed the same for all genotypes) and I take $\tau = (\mu_{2c} - \mu_0)/\sigma_E$. Values for τ were chosen on the basis of calculations given as supplemental material using

TABLE 1
Description of parameters

Parameter	Meaning	Default value
L	Number of segregating disease loci	—
π	Probability that an individual with no disease alleles is affected	0
π_L	$\pi_L = \pi^{1/L}$	0
δ	Probability that an individual with L disease alleles is affected	1
δ_L	$\delta_L = \delta^{1/L}$	1
h	Dominance coefficient for disease	0.5 (no dominance)
h_E	Dominance coefficient for expression-level trait	0.5
c	Number of disease loci influencing expression-level trait	1
M	Number of genes with expression levels influenced by disease locus	1
K	Disease prevalence	0.01
μ_0, μ_{2c}	Mean expression levels for 0 and $2c$ disease alleles, respectively	0, 1
σ_E	Genotype-specific expression-level standard deviation	0.333
τ	$\frac{\mu_{2c} - \mu_0}{\sigma}$	3
—	Number of genes tested on microarray	20,000

the data of HEDENFALK *et al.* (2001), who measured expression-level differences between patients carrying different breast cancer mutations. Briefly, I calculated t -statistics for differences in expression in 3326 genes between women carrying and not carrying BRCA1 mutations. I used an approach similar to that used by BREM and KRUGLYAK (2005) for correcting for multiple testing. That is, I randomly divided the data set in two and used half of the data to calculate t -statistics for all 3326 genes. Then, I used the other half of the data to recalculate the t -statistics for those genes with the 50 highest values. The 10th highest t -statistic from this set was ~ 3 (the maximum was 6 and the median was 2). On the basis of this, I take a baseline of $\tau = 3$. This was chosen as an extreme but not the most extreme value: that is, we are specifically interested in the expression-level traits that are most correlated with the disease mutation, but want to avoid possible outliers. See the supplemental material for more details. Table 1 gives a description of parameters used in this article.

Calculations of statistical power: Statistical power is defined as the probability of rejecting the null hypothesis given that it is false. In this case, the null hypothesis is that there is no difference in mean expression level between affected and unaffected individuals for the gene in question. The alternate hypothesis is that there is a difference. The goal here is to calculate, for a given set of parameter values, the power to detect expression-level difference for at least one gene whose expression-level trait is controlled by the TCL. I assume a typical setup for microarray experiments. There are n_1 affected individuals and n_2 unaffected individuals. Expression-level traits are measured for a large number of genes to find ones that are differentially expressed between the two treatments. The population means and variances for affected and unaffected individuals are calculated from (15). For sufficiently large sample sizes, the treatment means will have a normal distribution with these

means and variances. Using standard theory, the power to detect differences between treatments can be calculated. I assume a Bonferroni correction for multiple testing. Note that in standard false discovery rate approaches (BENJAMINI and HOCHBERG 1995) no tests are rejected unless the most significant one reaches the Bonferroni-corrected significance threshold. Thus, there is no difference between a Bonferroni correction and a false discovery rate approach because I am calculating the power to detect *any* differences.

In the case of reactive ELTs, I assume that the M expression-level traits are independent given the genotype at their causative loci. This is most likely not true in many cases. This assumption will increase the calculated power because correlations between ELTs would decrease the effective value of M .

Note that I consider detection of an association between a reactive ELT and the disease as a “success,” because that reactive ELT can be used to map the controlling disease locus even if the ELT itself is not disease causative. However, in another context a researcher might be interested in detecting only ELTs that are disease causative. In this case, the statistical power calculated here for reactive ELTs would actually be the probability of type I error.

In this article I strictly focus on expression-level traits directly controlled by disease-causative loci. There may also be expression levels whose controlling locus is linked to a disease locus. These expression levels will also be correlated with the disease, but this correlation will be weakened by recombination. To avoid complications, I do not consider such genes.

I use a simulation approach to calculate approximate power for $M > 1$. For each repetition, expression levels for treatment and control groups are randomly generated using distributions (1) and (2). A t -test is conducted for each of the M expression-level traits. If at least one expression-level trait is significantly different between

FIGURE 1.—Plot of power *vs.* L (number of disease loci) and h (the dominance coefficient). The power is the probability of rejecting the null hypothesis of no difference in expression between the treatments. Parameter values are $\pi = 0$, $\delta = 1$, $K = 0.01$, and $\sigma_g = 0.333$, and the sample size is 100 for each treatment. In both plots $\mu_0 = 0$ and $\mu_2 = 1.0$, where μ_i is the mean expression level for genotype with i disease alleles at the TCL. The dominance coefficient for expression h_E is 0 in a and 0.5 in b. The number of genes tested on the microarray was assumed to be 20,000.

treatment and controls, then that repetition is declared a success. The power is then approximately equal to the fraction of success in a large number of trials.

POWER TO DETECT EXPRESSION LEVEL DIFFERENCES DUE TO POLYMORPHISMS BETWEEN AFFECTED AND UNAFFECTED INDIVIDUALS: RESULTS

Reactive expression-level traits—multiplicative model: *Dependence on h and L :* Figure 1 shows surface plots of power to detect expression-level trait differences *vs.* L (effective number of causative loci) and h (dominance coefficient) for a multiplicative model. For each plot, disease prevalence $K = 0.01$, expression-level trait standard deviation $\sigma_E = 0.333$, sample size is 200 (100 for each treatment), the penetrance parameters π and δ are 0.0 and 1, respectively (indicating perfect penetrance with respect to the full multilocus genotype), and the expression-level trait depends only on a single TCL. Thus, there are three expression-level trait means. In both plots $\mu_0 = 0$ and $\mu_2 = 1.0$, where μ_i is

the mean expression level for individuals with i disease alleles at the TCL. These values correspond to $\tau = 3$. The dominance in expression is $h_E = 0$ (completely recessive) in Figure 1a and $h_E = 0.5$ (no dominance) in Figure 1b. I assume that there are 20,000 genes tested on the microarray. These are default parameter values (with $h_E = 0.5$) throughout this article unless otherwise specified.

The determinant of power is the amount of contrast between the treatment means. In Figure 1a, the expression-level trait is recessive and there is no mean difference between individuals with dd and Dd genotype at the causative locus. When $h = 0$ (disease risk is recessive with respect to the TCL), then all affected individuals are DD at the TCL. When $L = 1$, then all unaffected individuals are dd or Dd at the TCL and there is a strong contrast in expression level between the affected and unaffected treatments. The power is near 1. When h moves away from 0, then some Dd individuals are also affected. Because the disease allele is rare when $L = 1$, most affected individuals are Dd when $h > 0$ and thus there is little contrast between treatments. Thus, we see that the power quickly drops to near 0 as h is increased for $L = 1$.

When $L > 1$, then some nondisease individuals will also have DD genotypes at the TCL (see the description of the multiplicative trait model). As L is increased, the disease-allele frequency at the TCL also increases. As the number of causative loci is increased, then the probability of the disease events associated with those loci must also increase to maintain the constraint that the disease prevalence is 0.01. For example, the disease-allele frequency is 0.68 for $L = 6$ and 0.75 for $L = 10$ at $h = 0$. See SCHLIEKELMAN and SLATKIN (2002) for further details and plots of allele frequencies. As L increases, the proportion of unaffected individuals with a DD genotype at the TCL also increases. Thus, the contrast between the two treatments also decreases and power drops quickly. At $h = 0.1$, power is near 97% for $L = 6$, but drops to $\sim 46\%$ for $L = 11$. The power also drops quickly as h is increased from 0. For example, at $h = 0.3$, the power is $< 50\%$ for all values of L and for $h \geq 0.5$ the maximum power is ~ 0.05 . One interesting aspect of Figure 1a is that power actually increases when L is increased from 1 when h is low but > 0 . This occurs because allele frequency increases as L is increased. As discussed above, the disease-allele frequency is low for $L = 1$ and when $h > 0$ most affected individuals are genotype Dd . The allele frequency increases as L does and thus more affected individuals are of genotype DD . Only genotype DD produces any contrast in expression level, so power is increased.

Figure 1b has identical parameter values except that $h_E = 0.5$ (no dominance in expression). In this case, the power to detect differential expression is less sensitive to h and does not drop to zero as h is increased. Power is lowest for intermediate h and as above is highest for

FIGURE 2.—Plot of power *vs.* *c* (number of loci controlling expression-level trait). The y-axis is the probability of detecting a significant expression-level difference between affected and unaffecteds. The different curves correspond to the value of *L* as shown in a (the order is the same for b–d). The power is the probability of rejecting the null hypothesis of no difference in expression between the treatments. In a and b the expression-level trait has an additive dependence on genotype, while there is a multiplicative dependence in c and d. The number of genes tested on the microarray was assumed to be 20,000. Parameter values are $\pi = 0$, $\delta = 1$, $L = 9$, $K = 0.01$, $\sigma_g = 0.333$, $\mu_0 = 0$, $h_e = 0.5$, and $\mu_{2c} = 1$. The parameter $h = 0.5$ in a and c and $h = 0.9$ in b and d. The sample size is 100 for each treatment.

$h = 0$. However, the power in the best case is substantially worse for larger L than in the previous example. The maximum power is 0.81 for $L = 6$ and 0.61 for $L = 5$. As discussed above, disease-allele frequency is quite high for larger L when $h = 0$. Thus, the difference between treatment groups is primarily due to differences between Dd and DD individuals. Because the expression-level means are closer together, the difference is less. The closer h is to 0.5, the more affected individuals are of Dd genotype. Again, the contrast between treatments is lower and power decreases.

Supplemental material available at my website shows a similar plot for the case with $h_e = 1$. In this case the power is highest at $h = 1$ and drops off quickly as h is decreased from 1 (although less quickly than in Figure 1a).

Dependence on *c*: Next, I explore how the number of loci (*c*) controlling the expression-level trait affects the power to detect an expression-level difference. The more similar the genotypic dependence of the expression-level trait is to that of the disease, the more contrast in expression-level trait we expect there to be between affecteds and unaffecteds. Thus, increasing the number of loci controlling the expression-level trait should increase the power (since all loci controlling the expression-level trait are assumed to be controlling disease status also). There is no question that increasing the range of variation in an expression-level trait will increase power. The unknown is how changing the control of expression level affects power. Therefore, the maximum and minimum expression levels are kept constant as *c* is varied. That is, the expression level for the genotype with all *c* loci as DD is constant and the expression level for the genotype with all *c* loci as dd is constant. Figure 2 shows plots for several parameter values. In Figure 2, a and b, the expression-level trait has an additive dependence on the TCLs while it has a multiplicative dependence in

Figure 2, c and d. In all cases the disease has a multiplicative dependence. We see that the power increases only slowly with *c* in the additive case. The advantage of increasing *c* is that it increases the concordance of genotypic dependence between expression-level trait and disease status. However, if disease and expression-level traits depend in fundamentally different ways on the multilocus genotype, this concordance does not increase substantially with *c*.

In Figure 2, c and d, we see the unexpected result that the power initially increases with *c* but then decreases. The increase is because of the increased concordance of control of disease and genotype. That is, it is quite likely that an unaffected individual will have one or two disease alleles at a given locus (just not at all loci). Thus, if $c = 1$ then the ELT will often be differently regulated in unaffected individuals. It is less likely that the unaffected individual will have disease alleles at all of some set of multiple loci. Thus, the ELT is less likely to be differently regulated in unaffecteds when $c > 1$ and thus there is stronger association between disease and ELT and power initially increases with *c* (except for $L = 3$). The decrease occurs because of how the disease and expression models are defined. As *c* increases, it becomes likely that multiple controlling loci for the ELT in affecteds will be heterozygotic. In Figure 2c, the contributions to the expression means for dd , Dd , and DD individuals are 0, 0.5, and 1, respectively. Because these values are multiplied to get the mean, then the mean moves much closer to 0 than to 1 if there are more than one to two heterozygote loci. Thus, when *c* is large, then even affected individuals often have low expression means and this diminishes the contrast in expression between affected and unaffected. In Figure 2d, h is increased from 0.5 to 0.9. This leads to a decreased disease allele frequency (for constant K) and thus more het-

FIGURE 3.—Effect of M . The z-axis is the probability of detecting a significant expression-level difference between affecteds and unaffecteds for at least one of the M expression-level traits plotted against the parameters c and M . Parameter values are otherwise the same as those in Figure 2. The power calculations are based on simulations with 500 repetitions.

erozygote loci for affecteds. This in turn leads to more rapid loss of power as c increases. It is difficult to say whether this behavior is reasonable or should be considered an artifact of the model.

We see in Figure 2, c and d, that there is good power to detect expression-level differences for $L = 3$ and $c = 1$ –2 and for $L = 6$ cases and $c = 2$ –3. For higher values of L and $h = 0.5$ (Figure 2c) the power is maximized in the 50–60% range for $c = 4$ –5. For the $h = 0.9$ case (Figure 2d) power is substantially lower.

Dependence on M (number of expression-level traits controlled by the TCL): If the difference in genotype distribution between affecteds and unaffecteds means is not large at the TCLs, then the probability of the sample means being different enough to achieve significance may be small. In this case, the probability of achieving significance is increased if there are multiple expression-level traits dependent on the TCLs. Figure 3 shows a surface plot of power *vs.* c and M for $L = 9$, $h_E = h = 0.5$, and a sample size of 200. The power here is defined as the probability that at least one of the M ELTs is detected as differentially expressed. When $c = 1$, we see the power increasing from 5% for $M = 1$ to 50% for $M = 40$. In this case, there is little distinction between affecteds and unaffecteds at the TCL. The disease-allele frequency is 0.6 for these parameter values. Approximately 84% (that is, $1 - 0.4^2$) of unaffecteds carry a disease allele at the TCL and there is little difference in TCL genotype between affecteds and unaffecteds. With so little difference in genotype distribution, each additional ELT dependent on the TCL makes only a small difference. However, a large number of such ELTs do make a substantial difference. For the case $c = 2$, small changes in M have a substantial affect on power. When we jointly consider two loci, the genotypic difference between treatments is greater. Power is still poor (36%) for $M = 1$, but improves to >80% for $M = 8$. For $c = 3$, power

FIGURE 4.—Plot of power *vs.* sample size. The y-axis is the probability of detecting a significant expression-level difference between affecteds and unaffecteds. The different curves in a correspond to different values of c as shown. The curves in b correspond to values of M (first number) and c (second number). Parameter values are otherwise the same as those in Figure 2. The power curves in b are based on simulations with 100 repetitions. The waviness in the curves occurs because of the resulting sampling error in the simulations.

increases from ~60 to ~90% when M is increased from 1 to 4. Thus, increased M can help power substantially in cases where the genotypic distinction between treatments is intermediate.

Dependence on sample size: Figure 4 shows plots of power *vs.* sample size for various values of M and c for $L = 9$. $M = 1$ for all curves in Figure 4a, but varies in Figure 4b as shown. When $M = 1$, there is no case for which the power reaches 80% for sample sizes <250. A sample size of 600 is required for 80% power when $c = 1$. The best case is 80% power for a sample size of 250 when $c = 3$. The sample sizes become more reasonable (although still rather large) for higher M . Power is 85% and 92%, respectively, for $c = 3$ and $M = 4$ and 8 and a sample size of 160. However, a sample size of 350 is required to get 80% power even when $M = 8$ for $c = 1$ or 8.

Dependence on penetrance parameters: All of the above plots had perfect penetrance for the full multilocus genotype (that is, $\pi = 0$ and $\delta = 1$). Figure 5 shows that varying π (the probability of an individual with no disease alleles being affected) from zero seriously decreases power, while decreasing δ from 1 has negligible effect. For $\pi \geq 0.00005$, the power is near zero even for a sample size of 1500. When $\pi = 0$ it is certain that an affected individual has at least one disease allele at each locus. When $\pi > 0$, then some affected individuals do not have any disease alleles at some loci. For the parameters in Figure 5 ($L = 9$ and prevalence = 0.01) and $\pi = 0.0005$, it can be shown that most affected individuals have at least one locus with no disease alleles. Thus, the contrast in expression levels between affecteds and unaffecteds is diminished. On the other hand, many loci

FIGURE 5.—The effect of penetrance parameters. The y-axis is the probability of detecting a significant expression-level difference between affecteds and unaffecteds. The curves correspond to different values of the penetrance parameters π and δ as shown. $\pi = 0$ and $\delta = 1$ except when otherwise indicated. Parameter values are otherwise the same as those in Figure 2.

in unaffected individuals have a disease alleles regardless of the value of δ and changing it therefore has less effect. The sharp drop in power as π_L is increased is largely due to the particular form of model used. The parameter π is the probability of an individual with no disease alleles being affected. The parameter π_L , the contribution to disease risk of the TCL, is $\pi_L = \pi^{1/L}$. When L is large, π_L increases very quickly as π does. For example, when $L = 9$ and $\pi = 0.0005$, then $\pi_L = 0.0005^{1/9} = 0.43$. Thus, the probability that a locus with no disease alleles contributes its disease event is quite high. The effect of varying δ is small for the same reason. For $\delta = 0.5$ and $L = 9$, the value of δ_L is $0.5^{1/9} = 0.93$. It is, of course, unknown whether this form for π_L is appropriate. However, if we assume a multiplicative model and that individuals with no disease alleles can have the disease, then it is necessary.

Dependence on the heritability of expression-level trait: As the heritability of the expression-level trait with respect to the TCL increases, so does the power to detect that TCL. In the above results, I assumed $(\mu_{2c} - \mu_0)/\sigma = 3$, where μ_{2c} and μ_0 are the expression-level means for individuals with all $2c$ disease alleles and 0 disease alleles, respectively, and σ is the standard deviation in gene expression. This corresponds to 45% of variance explained by genotype. Referring to supplemental Figure 1, we see that a value of three is among the highest values calculated for the Hedenfalk data. However, there are genes with values ranging as high as six, and the power to detect expression level differences for these genes will be greater. Figure 6 shows a plot of power *vs.* the parameter $\tau = (\mu_{2c} - \mu_0)/\sigma$ for various values of L and c . In this figure μ_{2c} and μ_0 are held constant at 1 and 0, respectively, while the expression-

FIGURE 6.—The effect of the heritability of expression level. The y-axis is the probability of detecting a significant expression-level difference between affecteds and unaffecteds. The x-axis is the value of the parameter $\tau = (\mu_{2c} - \mu_0)/\sigma$. The curves correspond to different values L and c as shown. Parameter values are otherwise the same as those in Figure 2.

level variance σ is varied. The sample size is 200, $h = 5$, $h_E = 0.5$, and all other parameters are as in Figure 2. As expected, the power increases steadily with τ . For example, the power in the $L = 6$, $c = 1$ case increases from $\sim 7\%$ at $\tau = 2$ to 86% for $\tau = 8$. On the other hand, the power in the $L = 9$, $c = 1$ case reached only 37% at $\tau = 8$ because it started from only 1%.

Natural population—additive model: Power under an additive penetrance model is substantially worse than that under a multiplicative model. Figure 7a shows a plot of power to detect expression-level differences *vs.*

FIGURE 7.—Power for the additive penetrance model. The y-axis is the probability of detecting a significant expression-level difference between affecteds and unaffecteds with an additive penetrance model. The x-axis in a is the sample size and the x-axis in b is the parameter M . The value of L is 4 in b and as shown in a. The sample size in b is 100 in each group. Parameter values are otherwise the same as those in Figure 2. The power curves in b are based on simulations and are approximate. The waviness in the curves is due to the sampling error in the simulations.

FIGURE 8.—Power for causative ELT. The y-axis is the probability of detecting a significant expression-level difference between affecteds and unaffecteds for a causative ELT as described in the text. The x-axis in a is the sample size and the x-axis in b is the threshold Q . The sample size in b is 100 in each group. Parameter values are otherwise the same as those in Figure 2.

sample size for $L = 4$ and $L = 6$ and a range of c values for each. The power is very poor. With just four disease loci, sample sizes of >800 are required to reach 80% power. When $L = 6$, power is just 10% with an 800-sample size.

Crucially, the power is not affected much by c . This is in contrast to the multiplicative model, where increasing c has a large effect on power. Under the additive model, the mean expression levels for affecteds and unaffecteds do not change as c is varied. A proof of this is given in the supplemental results. Power increases with c under a multiplicative model (for small c) because increasing c increases the concordance between expression-level trait and disease. Under an additive model, each locus contributes independently and increasing c does not increase the concordance.

Power under a multiplicative penetrance is much higher than that under additive penetrance even for $c = 1$. This is due to the very different genotype distributions under the two models. With multiplicative penetrance, affected individuals have at least one disease allele at all loci (or at most loci if the penetrance parameter $\pi > 0$). In contrast, only one locus need have a disease allele to cause disease with additive penetrance. With $L = 6$ and other parameters as in Figure 7a, $\sim 90\%$ of affected individuals carry only a single disease allele and most of the remaining 10% carry single disease alleles at two loci. Thus, only $(90\%/6 + 10\%/3) = 18\%$ of affected individuals carry a disease allele on a given disease locus. Assuming no dominance as in Figure 7, the average expression level for a gene controlled by this locus is 0.5 and thus the average expression level among affecteds is $0.5 \times 0.18 = 0.09$. Most unaffecteds carry no disease alleles, so the mean difference in expression level between affecteds and unaffecteds is ~ 0.09 . With the same parameters and a multiplicative model, affecteds on average have three loci that are homozygous for the disease allele and three loci that are heterozygous. The average expression level in affecteds for a gene controlled by the disease locus would be $\sim 0.5 \times 1 + 0.5 \times 0.5 = 0.75$. Now, as discussed earlier, disease-allele frequencies are high under a multiplicative model (in strong contrast to the case with an additive model). The

disease-allele frequency would be 0.46 for the parameters in Figure 7 and genotype frequencies in unaffecteds would be ~ 0.21 disease homozygote, 0.50 heterozygote, and 0.29 normal homozygote. Thus, the mean expression level in unaffecteds would be $0.21 \times 1 + 0.5 \times 0.5 + 0.29 \times 0 = 0.46$ and the mean difference between affecteds and unaffecteds would be $\sim 0.75 - 0.46 = 0.29$. The mean difference in expression level is 3.5 times higher for the multiplicative model than for the additive one and power to detect expression-level differences is correspondingly higher under the multiplicative model.

Figure 7b shows a plot of power *vs.* M for sample size 200, $L = 4$, and $c = 1, 3$, and 5. We see increasing M has minor impact. Even for this low value of L the power is still low for M up to 20.

Causative expression-level trait: There are many possibilities for how disease risk depends on the expression level of a causative transcript. I use a simple threshold model with $u(X_1) = 1$ if $X_1 \geq Q$ and $u(X_1) = 0$ if $X_1 < Q$. The expression means and variances for affecteds and unaffected individuals are calculated using Equation 15 in a similar manner to that done for the reactive ELT. For example, the expected value among affecteds for the multiplicative model is

$$E(X_1 | T = 1) = \frac{K_{c+1} \dots K_L}{K} \times \sum_{g_1 \dots g_c} \left[\int_Q^\infty X_1 f(X_1 | \mu_{g_1 \dots g_c}, \sigma_{g_1 \dots g_c}) dX_1 \right] \times P(g_1 \dots g_c). \quad (17)$$

Figure 8a shows a plot of power to detect a difference in X_1 between affecteds and unaffecteds as a function of sample size. The parameter $Q = 0.5$, $c = 1$, and the expression means are $\mu_0 = 0$, $\mu_1 = 0.5$, and $\mu_2 = 1$. The different curves correspond to the value of L as shown. Other parameters are as in Figure 2. The power is greatly improved compared to a reactive ELT. For example, we see that 80% power is achieved at a sample size of ~ 100 when $L = 9$. This compares to a sample size of 600 for

80% power for similar parameter values with a reactive ELT. This increased power results from a combination of higher expression means and lower variances for affected individuals when the ELT is causative relative to the case when it is reactive. Both of these effects result primarily from the heterozygote expression distribution being truncated in affected individuals. The value of the ELT must exceed Q in affected individuals. With the parameter values in Figure 8 ($Q = 0.5$ and $\mu_1 = 0.5$), half of the expression distribution is truncated in affected heterozygotes.

Figure 8b shows the effect of the threshold Q . We see that for these parameter values the power is at a peak at $\sim Q = 0.6$, but that Q has a small impact on power. The effects of varying Q are complicated. As Q increases, fewer individuals will have an ELT value that exceeds it. Thus, the disease-allele frequency must increase to keep disease prevalence fixed. Increasing Q has four major effects. First, it increases the expression mean in affected individuals, which tends to increase power. Second, it increases the expression mean in unaffected individuals (because the disease-allele frequency increases), which tends to decrease power. Third, increasing Q reduces the portion of each genotype-specific expression distribution that exceeds the threshold, tending to decrease variance. Fourth, when Q reaches extreme values, then the disease allele frequency does also. This tends to reduce expression variance by reducing genetic variance in the population. The interaction between these effects is complex, but the net effect on power is small because the various effects are often in opposition to each other.

Figures available as supplemental material show the effects of various other parameters on the power for the causative ELT. Power increases very substantially with c as the proportion of the risk explained by the ELT increases. The degree of dominance in the ELT (that is, the mean expression level for heterozygotes) has a roughly inverse effect to that of Q . That is, power is at a peak for intermediate dominance and the various effects discussed above for Q work in roughly the opposite direction as h is increased.

Allowing M to vary for causative ELTs would require specifying a model for how these ELTs interact to determine disease risk. Thus, I keep M fixed at one. Other parameters have effects similar to those observed for reactive ELTs.

This model for a causative ELT is not compatible with additive disease risk for the standard parameter values used in this article. In the additive case, the disease risk associated with the causative locus is $K/L = 0.01/9 = 0.0011$. Under the standard parameters, the ELT has a probability of 0.07 of exceeding the threshold $Q = 0.5$ when its controlling locus has no disease alleles. This means that an individual with no disease alleles still has a probably 0.07 of being affected, which far exceeds the "allowed" disease risk for the locus. This can be corrected by lowering the genotype-specific ELT standard

deviation σ_E to a level that makes the probability of exceeding $Q < 0.011$. This requires halving the standard deviation to ~ 0.15 for the standard parameter values. Not surprisingly, power is quite good with this low ELT variance (comparable to that seen in Figure 8). However, the analysis of σ_E values described in the METHODS section indicates that ELTs with values this low may be very rare.

POWER TO MAP CAUSATIVE LOCI USING DIFFERENTIALLY EXPRESSED TRANSCRIPTS AS QUANTITATIVE TRAITS

Above, I considered the power to detect differences in expression-level traits between disease-affected and -unaffected individuals. For causative ELTs, good power to detect association with the disease is achieved with $c = 1$ and sample sizes of ~ 100 . With $c = 1$ (meaning that the ELT has a single controlling locus), the power to map the controlling locus will be good. Thus, in this case, eQTL-based strategies for genome scans should be powerful.

The situation is more complicated for reactive ELTs. Power was shown to be very poor if an additive penetrance model holds. Power for multiplicative penetrance depends strongly on the parameter c . If there is only a single TCL controlling an expression-level trait ($c = 1$), then power to detect an association between that gene and disease status is low unless L is small. If multiple TCL control the gene (*i.e.*, $c > 1$), then the power to detect associations with the disease is higher. However, this expression-level trait is then a multilocus trait and it will have lower power in use as a quantitative trait for detecting the TCL(s).

A number of methods for mapping QTL in natural populations have been proposed (see, *e.g.*, LYNCH and WALSH 1998). I will focus on humans and assume that mapping is conducted using Haseman–Elston (HE) regression (HASEMAN and ELSTON 1972). HE regression uses sib pairs and detects linkage to a QTL by regression of the squared difference in trait values on the number of alleles identical-by-descent (IBD) between the sibs at each marker. If a marker is linked to a QTL, then there should be a negative correlation between the squared trait difference and the IBD status at the marker.

I used a Monte Carlo simulation to estimate power for HE regression with an eQTL. I assumed a multiplicative model for gene expression with equally contributing loci. Figure 9a shows a plot of power *vs.* the parameter c for sample sizes of 100, 300, and 500 sib pairs. $M = 1$ in Figure 9a. That is, there is one expression-level trait controlled by the eQTL. This expression-level trait being tested is assumed to be completely linked to one of the c controlling loci. A Bonferroni correction for 50 tests was applied (*i.e.*, $\alpha = 0.05/50 = 0.001$). This is very a modest correction considering that there typically might be tens or hundreds of differentially expressed genes tested

FIGURE 9.—Power to map eQTL using Haseman-Elston regression. Power was calculated using a Monte Carlo simulation with 500 replicates. The marker being tested is assumed to be completely linked to one of the c controlling loci for the expression-level trait. The y-axis shows the power to detect this marker as being linked to the expression-level trait. In a, the parameter $M = 1$ and the sample size is shown for each curve. In b, the sample size is 500 sib pairs and the value of M is shown for each curve. A significance threshold of $\alpha = 0.05/50 = 0.001$ was applied. Parameter values are otherwise the same as those in Figure 2.

against each of hundreds of markers (note that a correction for 20,000 tests was used in previous sections of this article). We see that power drops quickly as c increases, with $<10\%$ power for $c = 2$ and 500 sib pairs.

The power to detect a causative locus is increased if it is an eQTL for multiple expression-level traits. That is, if there are M expression-level traits controlled by the locus then there are M chances to detect it. However, these are not independent chances because the sample individuals are the same for each expression-level test. Thus, genotypes are not replicated. The expression-level traits are only independent conditioned on genotype.

The power curves in Figure 9b correspond to $M = 10, 20, 50$, and 100 as shown. The sample size is 500 sib pairs in each case. The power is defined as the probability that the target causative locus is detected as an eQTL for at least one of the expression-level traits. Each of the M expression-level traits is assumed to be controlled by the same set of c loci. We see that the power increases strongly with M . However, $M = 100$ is required to achieve 80% power when $c = 2$ and the power is only 20% when $c = 3$.

The basic HE regression method assumed here is not the most powerful method available for detecting QTL in natural populations. However, other methods are not dramatically better and it is not expected that the qualitative results will be much different. Thus, we see that the power for detecting an eQTL is expected to be poor when the expression-level traits have more than one controlling locus, unless that locus is an eQTL for many expression-level traits.

DISCUSSION

There are too many unknowns to draw any firm conclusions about the likely effectiveness of mapping complex trait loci using gene expression levels as quantitative traits. There is strong evidence that a substantial number of gene expression-level traits in many organisms have a relatively simple genetic basis and high heritability relative to that seen for complex clinical traits. Furthermore, SCHADT *et al.* (2003), BRYSTRYKH *et al.* (2005), CHESSLER *et al.* (2005), HUBNER *et al.* (2005), and

others have identified probable associations of eQTL with clinical traits in mice or rats. Expression-level traits with high heritability will give good power for mapping their associated eQTL. However, I have shown here that establishing that these eQTL are also associated with the trait of interest may be difficult. If such expression-level traits are causative, in the sense of disease risk depending directly on expression level, then sample sizes on the order of 100 are sufficient for good power. On the other hand, if the ELT is reactive to a causative locus, then there is a trade-off between the power to show an association between expression-level trait and disease and the power to map eQTL for that expression-level trait. The simpler the genotypic dependence of an expression-level trait is on a disease-causative eQTL, the easier it is to map that eQTL. From a mapping perspective, the ideal case is that the expression-level trait depends on a single causative locus. However, in this case the correlation with the disease will be small. Most of the results in this article with $c = 1$ and $L > 3$ for a reactive ELT show poor power to detect differential expression between affecteds and unaffecteds unless sample size is well over 500 individuals. This power increases quickly as c does, although the sample sizes required for 80% power are all rather high compared to most current microarray experiments. Unfortunately, the power to map causative loci drops quickly with c . Figure 9 shows that for Haseman-Elston regression the power to detect linkage is $<10\%$ for $c > 1$ even with 500 sib pairs. This indicates that for humans there may be no expression-level traits that give good power both to detect reactive expression-level differences between affecteds and unaffecteds and to map their underlying causative loci. The underlying problem is that power is poor for mapping QTL in natural populations. This is somewhat ameliorated if the disease locus is an eQTL for many expression-level traits. However, there must be on the order of hundreds of controlled expression-level traits before power becomes reasonable. I stress again, however, that these power calculations are not meant to be taken too literally because there are many other possible methods for doing genome scans using eQTL. Thus, the results are intended more as a guide to intuition.

Results that will appear in a future publication (P. SCHLIEKELMAN, unpublished results) show that the situation is better for crosses with inbred lines. The power to detect expression-level differences is not substantially higher, but the power to detect linkage drops off more slowly as the number of loci (c) controlling the expression-level trait increases.

These results do not imply that no reactive expression-level changes should be detected with natural populations. On the contrary, the power to detect such differences is quite reasonable for expression-level traits influenced by multiple causative loci (*i.e.*, expression-level traits with $c > 1$). The point is that the reactive expression-level traits that are found to be most differentially expressed are likely to be poor candidates for use as quantitative traits for linkage detection.

Model assumptions: The results of this article depend on a large number of assumptions about how disease risk and expression levels depend on genotype. This is unavoidable given our lack of knowledge of the genetics of complex traits. I have used two very different disease-penetrance models and a wide range of parameter values to give a sense of what the range of possibilities is. However, in the end the results of a model are only as good as the assumptions that go into it and it is possible that these assumptions are badly off.

In a sense, the additive and multiplicative models are at extreme ends of a continuum. A multiplicative model requires that each of the causative loci have a disease event for the disease to occur. The additive model requires only a single locus to have a disease event for disease to occur. As shown in this article, this difference results in very different genotype distributions. I have focused on the multiplicative model because the power to detect reactive expression-level differences is very poor under an additive model and the additive model does not appear consistent with a causative ELT. If an additive model is correct, then there appears to be little hope for mapping complex disease genes using eQTL. Fortunately, additive disease-penetrance models do not appear to be consistent with family history data in humans. RISCH (1990) showed that inheritance patterns do not change as the number of causative loci increases under an additive model. Thus, if the inheritance patterns for a disease indicate a multilocus trait, then it cannot be additive. Multiplicative models typically do fit family history data well (*e.g.*, SCHLIEKELMAN and SLATKIN 2002). However, it is likely that a great many other models would fit equally well and there is no way to tell if the “true” model behaves more like an additive model, like a multiplicative model, or like neither model.

The values of M (the number of conditionally independent expression-level traits influenced by the TCL) assumed here are also highly speculative. Surveys of eQTL have found great variability in the number of gene expression-level traits mapping to individual loci. SCHADT *et al.* (2003) found several hotspots with hun-

dreds of eQTL mapping to 4-cM regions. CHESLER *et al.* (2005) found a single locus modulating the expression of 1650 transcripts in addition to numerous loci modulating hundreds of transcripts. On the other hand, MONKS *et al.* (2004) did not find evidence of such hotspots. Their study was conducted with randomly chosen human families whereas studies that have found hotspots were conducted with inbred lines. One possible explanation for the difference is that there is little variation at hotspot loci in natural populations, but selection for divergent phenotypes (or perhaps chance) in inbred lines leads to genotypic differences between strains. Another possibility is that expression-level traits with multilocus inheritance in natural populations become monogenic between inbred lines and thus are easier to detect. In any event, an increase in M increases the power to detect at least one expression-level trait associated with the TCL and a causative locus with very high M would be detectable as differentially expressed with reasonable sample sizes even for $c = 1$ (see Figure 3). However, M is the effective number of *independent* (given TCL genotype) expression-level traits influenced by the TCL. The amount of correlation between expression-level traits in the above studies is unknown. Most likely they are not all independent and the effective number of independent expression-level traits may be substantially lower than the observed number.

Comparisons to standard methods for genome scans: The required sample sizes shown here are large by the standards of typical microarray experiments, but not outrageous on the scale of human linkage studies. Power studies (*e.g.*, RISCH and MERIKANGAS 1996; SLAGER *et al.* 2000) have shown that sample sizes of many thousands might be required to detect complex trait loci in genome scans. There are too many unknowns to say whether eQTL will provide a more effective method for mapping complex trait loci than do conventional mapping strategies. If there are causative ELTs for many complex traits, then the results here indicate that eQTL-based strategies may be very effective. The central point of this article with respect to reactive eQTL is that high power in mapping of eQTL is traded for low power in correlating the eQTL with traits of interest. It remains to be determined exactly how low. If we assume that transcripts with high heritability relative to causative loci (and thus linkage or association is easy to detect) exist, then the fundamental question is how the power to detect correlations between expression-level traits and disease compares to the power to detect linkage or association between the disease and the causative loci. There is reason for optimism. If there are typically many genes with expression levels mapping to single eQTL (*i.e.*, high M) and if τ -values are typically very high, then required sample sizes drop down to the range of ≤ 200 . Furthermore, the experimental design and method of data analysis that I have assumed can certainly be improved (see the next section).

The results here show that the power to detect association of a causative ELT with a disease locus compares very favorably with the power to detect complex trait loci using traditional mapping methods. Why should this be so? That is, the locus controlling the ELT in Figure 8 accounts for one-ninth of the disease risk. Numerous gene-mapping power studies show that the power to detect such a locus is poor. Why is the power apparently better to detect disease association with an ELT that accounts only for one-ninth of the risk? The answer to this can be understood by analogy with association studies for complex traits. It is well established (*e.g.*, RISCH and MERIKANGAS 1996) that association studies are powerful for detecting common disease alleles when there are only two alleles in the population. However, power in association studies plummets when there are multiple alleles (SLAGER *et al.* 2000). In detecting association of a causative eQTL with a complex trait, the ELT is a proxy for genotype. If the genotype-specific expression means were widely enough separated, then we could unambiguously assign individuals to groups corresponding to genotype and conduct an association test directly. We do a similar thing when we do a *t*-test of an ELT determined by a mixture of expression distributions. We will tend to find a significant difference in expression between affecteds and unaffecteds if the allele frequency at the locus controlling the transcript is substantially different between affecteds and unaffecteds, which is the same requirement for an association test. It can be shown (P. SCHLIEKELMAN, unpublished data) that there are strong similarities between the functional forms of the noncentrality parameters that determine power for a case-control test and for a *t*-test of an ELT determined by a mixture of genotype-specific expression distributions and that these tests have similar dependence on population genetic parameters. A key advantage of an ELT approach is that there are effectively only two allele types. That is, all alleles that increase expression (and therefore disease risk in this study) are collapsed into a single disease-risk-increasing group and all alleles that decrease expression are grouped into a risk-decreasing group. Thus, testing for an association between an ELT and a clinical trait has a strong similarity to association testing with just two alleles.

A key advantage of ELT-based methods is that they identify direct correlations to genes and not to markers. Even if the identified gene is reactive only to a disease locus and not disease causative, there is still no recombination with a marker to diminish power. It appears that establishing association between transcripts and trait will be the low-power stage in genome scans using eQTL, while establishing linkage between a transcript and a marker will be relatively easy. Having no recombination involved in the low-power stage may be a major advantage for eQTL genome scans. Of course, there will also likely be genes whose expression level is controlled by a locus

linked to a disease locus and where recombination is a factor. Such genes will also be correlated with the disease, but more weakly if all else is the same.

Improvements in power: There appears to be substantial potential for increasing the power of genome scans using eQTL beyond what I have shown here. SCHADT *et al.* (2003) stressed the importance of combining information from all transcripts instead of focusing on transcripts individually as I have done here. Several studies (GHAZALPOUR *et al.* 2005, 2006; LI *et al.* 2006) have used approaches that combine information from many transcripts in what are essentially genome scans. For example, GHAZALPOUR *et al.* (2006) identified 12 transcription modules using gene coexpression networks constructed from expression data for 3421 genes in an F₂ cross between inbred mice strains. They then looked for correlations between these transcription modules and obesity-related clinical traits by calculating the average correlation between the traits and genes in the module. Finally, they mapped eQTL for all genes within the modules and looked for genomic regions that were enriched for eQTL for each module. *cis*-eQTL in these regions were then taken as candidate genes for traits correlated with the corresponding transcription module. For example, they found one module significantly correlated with body weight and nine genomic regions enriched for this module. One of the major advantages of this approach relative to the assumptions of the analysis presented here is that the gene module approach greatly reduces the number of tests for correlation (that is, from thousands down to 12). This greatly increases the power to detect a correlation with the trait. A potential drawback is that there is no way to tell whether a particular candidate region is causative for the trait or is causative only for the module. Still, this type of approach seems very promising.

Experimental designs utilizing family structure will reduce the number of segregating loci between affected and unaffected subjects and this can be used to increase power for detecting expression-level differences associated with causative loci. KRAFT *et al.* (2003) proposed a family-based test for correlation between gene expression and a quantitative trait. They presented power simulations showing very good power for their proposed test, but did not model the genetics of the trait or expression level. Instead they assumed correlations of 33 or 67% between trait and expression level. The results here indicate that this assumed correlation is far too high for complex traits. While their method is promising, the power cannot be expected to be as high as their simulations show.

PEREZ-ENCISO *et al.* (2003) proposed a partial least-squares regression approach to construct "supergenes" composed of linear combinations of transcripts chosen to maximize correlation with the trait of interest. This linear combination of expression levels is then used as a quantitative trait for mapping QTL. Simulations showed

that this method gives substantial improvements in power over using the trait directly for QTL mapping. PEREZ-ENCISO *et al.* (2003) assumed a large QTL effect compared to the effects for causative loci shown in this article and as a result showed good power to detect the QTL for many scenarios. It is not clear that the increase in power for their method would be sufficient to make much difference if the QTL effect is small. The results of this article indicate that we might expect large groups of transcripts controlled by individual causative loci to have a weak association with the trait. This suggests that methods of analysis that use joint association of sets of genes with the disease might be an effective strategy.

I thank Mark Schliekelman and several anonymous reviewers for helpful comments on this manuscript.

LITERATURE CITED

- ALTMULLER, J., L. J. PALMER, G. FISCHER, H. SCHERB and M. WJST, 2001 Genomewide scans of complex human diseases: true linkage is hard to find. *Am. J. Hum. Genet.* **69**: 936–950.
- BENJAMINI, Y., and Y. HOCHBERG, 1995 Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**: 289–300.
- BERGE, K. E., H. TIAN, G. A. GRAF, L. YU, N. V. GRISHIN *et al.*, 2000 Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science* **290**: 1771–1775.
- BREM, R. B., and L. KRUGLYAK, 2005 The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. USA* **102**: 1572–1577.
- BREM, R. B., G. YVERT, R. CLINTON and L. KRUGLYAK, 2002 Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**: 752–755.
- BYSTRYKH, L., E. WEERSING, B. DONTJE, S. SUTTON, M. T. PLETCHER *et al.*, 2005 Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat. Genet.* **37**: 225–232.
- CHESLER, E. J., L. LU, S. SHOU, Y. QU, J. GU *et al.*, 2005 Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.* **37**: 233–242.
- CHEUNG, V. G., L. K. CONLIN, T. M. WEBER, M. ARCARO, K. Y. JEN *et al.*, 2003 Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.* **33**: 422–425.
- GHAZALPOUR, A., S. DOSS, S. S. SHETH, L. A. INGRAM-DRAKE, E. E. SCHADT *et al.*, 2005 Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol.* **6**: R59.
- GHAZALPOUR, A., S. DOSS, B. ZHANG, S. WANG, C. PLAISIER *et al.*, 2006 Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* **2**: 1182–1192.
- GIBSON, G., and B. WEIR, 2005 The quantitative genetics of transcription. *Trends Genet.* **21**: 616–623.
- GLAZIER, A. M., J. H. NADEAU and T. J. AITMAN, 2002 Finding genes that underlie complex traits. *Science* **298**: 2345–2349.
- HASEMAN, J. K., and R. C. ELSTON, 1972 The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**: 3–19.
- HEDENFALK, I., D. DUGGAN, Y. CHEN, M. RADMACHER, M. BITTNER *et al.*, 2001 Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344**: 539–548.
- HSIEH, W. P., G. PASSADOR-GURGEL, E. A. STONE and G. GIBSON, 2007 Mixture modeling of transcript abundance classes in natural populations. *Genome Biol.* **8**: R98.
- HUBNER, N., C. A. WALLACE, H. ZIMDAHL, E. PETRETTO, H. SCHULZ *et al.*, 2005 Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet.* **37**: 243–253.
- KARP, C. L., A. GRUPE, E. SCHADT, S. L. EWART, M. KEANE-MOORE *et al.*, 2000 Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat. Immunol.* **1**: 221–226.
- KENDZIORSKI, C., and P. WANG, 2006 A review of statistical methods for expression quantitative trait loci mapping. *Mamm. Genome* **17**: 509–517.
- KIRST, M., A. A. MYBURG, J. P. G. DE LEON, M. E. KIRST, J. SCOTT *et al.*, 2004 Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiol.* **135**: 2368–2378.
- KRAFT, P., E. SCHADT, J. ATEN and S. HORVATH, 2003 A family-based test for correlation between gene expression and trait values. *Am. J. Hum. Genet.* **72**: 1323–1330.
- LAWN, R. M., D. P. WADE, M. R. GARVIN, X. WANG, K. SCHWARTZ *et al.*, 1999 The Tangier disease gene product ABC1 controls the cellular apolipoprotein-mediated lipid removal pathway. *J. Clin. Invest.* **104**: R25–R31.
- LI, H., L. LU, K. F. MANLY, E. J. CHESLER, L. BAO *et al.*, 2005 Inferring gene transcriptional modulatory relations: a genetical genomics approach. *Hum. Mol. Genet.* **14**: 1119–1125.
- LI, H. Q., H. CHEN, L. BAO, K. F. MANLY, E. J. CHESLER *et al.*, 2006 Integrative genetic analysis of transcription modules: towards filling the gap between genetic loci and inherited traits. *Hum. Mol. Genet.* **15**: 481–492.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
- MONKS, S. A., A. LEONARDSON, H. ZHU, P. CUNDIFF, P. PIETRUSIAK *et al.*, 2004 Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75**: 1094–1105.
- MORLEY, M., C. M. MOLONY, T. M. WEBER, J. L. DEVLIN, K. G. EWENS *et al.*, 2004 Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**: 743–747.
- OLEKSIK, M. F., G. A. CHURCHILL and D. L. CRAWFORD, 2002 Variation in gene expression within and among natural populations. *Nat. Genet.* **32**: 261–266.
- PEREZ-ENCISO, M., M. A. TORO, M. TENENHAUS and D. GIANOLA, 2003 Combining gene expression and molecular marker information for mapping complex trait genes: a simulation study. *Genetics* **164**: 1597–1606.
- RISCH, N., 1990 Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.* **46**: 222–228.
- RISCH, N., and K. MERIKANGAS, 1996 The future of genetic studies of complex human diseases. *Science* **273**: 1516–1517.
- SCHADT, E. E., S. A. MONKS, T. A. DRAKE, A. J. LUSIS, N. CHE *et al.*, 2003 Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**: 297–302.
- SCHADT, E. E., J. LAMB, X. YANG, J. ZHU, S. EDWARDS *et al.*, 2005 An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**: 710–717.
- SCHLIEKELMAN, P., and M. SLATKIN, 2002 Multiplex relative risk and estimation of the number of loci underlying an inherited disease. *Am. J. Hum. Genet.* **71**: 1369–1385.
- SLAGER, S. L., J. HUANG and V. J. VIELAND, 2000 Effect of allelic heterogeneity on the power of the transmission disequilibrium test. *Genet. Epidemiol.* **18**: 143–156.
- WAYNE, M. L., and L. M. MCINTYRE, 2002 Combining mapping and arraying: an approach to candidate gene identification. *Proc. Natl. Acad. Sci. USA* **99**: 14903–14906.
- YAN, H., W. YUAN, V. E. VELCULESCU, B. VOGELSTEIN and K. W. KINZLER, 2002 Allelic variation in human gene expression. *Science* **297**: 1143.

Communicating editor: M. W. FELDMAN