An Overview of Statistical Approaches for Expression Trait Loci Mapping

Christina Kendziorski and Meng Chen

0.1 Introduction

Karl Sax was a pioneer in the field of quantitative trait loci (QTL) mapping. In his ground breaking 1923 paper, Sax identified a quantitative trait locus (QTL) for seed weight by associating the trait with seed color (a "marker" for which genotype information could be inferred). The next 60 years saw only a handful of similar studies, largely due to limitations imposed by the difficulty in arranging crosses with a reasonably large number of genetic markers. This changed in the 1980s following the discovery that abundant, highly polymorphic variation could be used to derive molecular markers densely spaced throughout the genome (Botstein *et al.* 1980). This advance, combined with statistical methods for QTL mapping (Lander and Botstein 1989), led to hundreds of QTL mapping studies.

A recent advance of comparable significance has been made in the area of phenotyping. With high throughput technologies now widely available, investigators today can easily measure thousands of traits for QTL mapping. Gene expression abundances measured via microarrays are particularly amenable to QTL mapping, and most scientists agree that the mapping of gene expression has the potential to impact a broad range of biological endeavors (Cox 2004; Broman 2005).

The optimism is based largely on the first expression trait loci (ETL) studies which have demonstrated utility in identifying candidate genes (Schadt *et al.* 2003; Bystrykh *et al.* 2005 Hubner *et al.* 2005), in inferring not only correlative but also causal relationships between modulator and modulated genes (Brem *et al.* 2002; Schadt *et al.* 2003; Yvert *et al.* 2003), in elucidating subclasses of clinical phenotypes (Schadt *et al.* 2003; Bystrykh *et al.* 2005; Chesler *et al.* 2005; Hubner *et al.* 2005), and perhaps most importantly, in identifying "hot spot" regions, genomic regions where multiple transcripts map (Schadt *et al.* 2003; Brem *et al.* 2002; Morley *et al.* 2004; Bystrykh *et al.* 2005; Chesler *et al.* 2005; Hubner *et al.* 2005). Hot spot regions are attractive for follow up studies as they putatively contain master regulators that affect transcripts of common function. The identification of master regulators could give critical information on mechanisms of regulation that remain poorly characterized and ultimately lead to targets of gene therapies (Cox 2004; Schadt *et al.* 2003). As a result of these successes, a number of efforts are now underway to localize the genetic basis of gene expression.

It is clear that the experimental set up in an ETL mapping study is structurally similar to a traditional QTL mapping study, but with thousands of phenotypes; and, as a result, most published studies to date have used methods developed for the QTL mapping problem in the ETL mapping setting. Lan *et al.* (2003) reduced the expression measurements to a few summary scores using a principal components analysis and then used single-trait QTL mapping methods to map the summary phenotypes. Doing so proved useful; however, transcript specific information could not be recovered. Others have used a "transcript-based" approach. In a transcript-based approach, each transcript is treated separately as a one-dimensional phenotype for QTL mapping. Single trait QTL analysis is then carried out thousands of times (once for each transcript). Notably, although adjustments are made for multiple tests across the genome,

no adjustments are made for multiple tests across transcripts. This leads to a potentially serious multiple testing problem and an inflated false discovery rate (FDR).

An alternative approach recognizes the similarities between ETL mapping and the problem of identifying differentially expressed (DE) transcripts in a standard microarray experiment. By grouping animals with similar marker genotypes, the ETL mapping problem at a particular marker reduces to identifying DE transcripts across the genotype groups. Any method developed for identifying DE transcripts could be applied. Similar to the transcript-based approach, this "marker-based" approach is also subject to inflated FDR as here multiplicities across markers are not accounted for. For some labs, an inflated FDR is tolerable as many genes can be tested quickly for certain properties and discarded if found to be false positives. However, for many labs, validation tests are prohibitively expensive and statistical methods that control error rates across both markers and transcripts are needed. Kendziorski *et al.* (2004) proposed such an approach, the mixture over markers (MOM) model.

In this chapter, we will review transcript-based approaches, marker-based approaches, and the MOM model approach to ETL mapping. The advantages and disadvantages of these approaches are discussed in Sections 0.2 and 0.4. Utility is evaluated using simulated data and data from two case studies (Section 0.3).

0.2 ETL Mapping Data and Methods

0.2.1 Data

The general data collected in an ETL mapping experiment consists minimally of a genetic map, marker genotypes, and microarray data (phenotypes) collected on a set of individuals. A genetic marker is a region of the genome of known, or estimated, location. These locations make up the genetic map. At each marker, genotypes are obtained. ETL mapping studies take place in both human and experimental populations. We focus here on the latter. For these populations, the number of possible marker genotypes is relatively small.

Studies with experimental populations most often involve arranging a cross between two inbred strains differing substantially in some trait of interest to produce F1 offspring. Segregating progeny are then typically derived from a B1 backcross (F1 x Parent) or an F_2 intercross (F1 x F1). Repeated intercrossing ($F_n x F_n$) can also be done to generate so-called recombinant inbred (RI) lines. For simplicity of notation, we focus on a backcross population. This is not required and is relaxed in the simulation and case studies sections. Consider two inbred parental populations P_1 and P_2, genotyped as AA and aa, respectively, at M markers. The offspring of the first generation (F_1) have genotype Aa at each marker (allele A from parent P_1 and a from parent P_2). In a backcross, the F_1 offspring are crossed back to a parental line, say P_1 resulting in a population with genotypes AA or Aa at a given marker. We denote AA by 0 and Aa by 1.

For each member of the backcross population, phenotypes are collected via microarrays. For the k^{th} animal, let $y_{t,k}$ denote the expression level for transcript t and $g_{m,k}$ denote the genotype at marker m; $t = 1, 2, \ldots, T$ and $k = 1, 2, \ldots, n$. To avoid confusion when referring to genes on a genetic map and gene expression levels measured on a microarray (where the physical location of the gene is often not known), when referring to the former, we use the term gene; when referring to the latter, we use transcript or trait.

Most questions addressed in an ETL mapping study rely on the ability to identify a list of significant linkages between transcripts and markers. To be precise, a transcript t is linked to marker m if $\mu_{t,0} \neq \mu_{t,1}$, where $\mu_{t,0(1)}$ denotes the latent mean level of expression of transcript t for the population of animals with genotype $0(1)$ at marker m. Suppose observations $y_{t,k}$ have density $f_{obs}(y_{t,k}|\mu_{t,g_{m,k}}, \theta)$ where θ denotes any remaining unknown parameters. Assuming independence across animals, under the null hypothesis of no linkage, the data is governed by $\prod_{k=1}^{n} f_{obs}(y_{t,k}|\mu_{t,0} = \mu_{t,1}, \theta)$; and under the alternative, $\prod_{k=1}^{n} \left[f_{obs}(y_{t,k}|\mu_{t,0}, \theta)\right]^{1-g_{m,k}} \left[f_{obs}(y_{t,k}|\mu_{t,1}, \theta)\right]^{g_{m,k}}$. As discussed below, a main difference between the transcript-based (TB) and marker-based (MB) approaches arises from different assumptions regarding the latent means.

0.2.2 Transcript Based Approach

A TB approach refers generally to the repeated application of any single phenotype QTL mapping method to each mRNA transcript, with locations identified as important if the test statistic of interest exceeds some critical value. The LOD score

$$\log_{10}\left(\frac{\prod_{k=1}^{n} f_{obs}(y_{t,k}|\hat{\mu}_{t,0}, \hat{\mu}_{t,1}, \hat{\theta})}{\prod_{k=1}^{n} f_{obs}(y_{t,k}|\hat{\mu}, \hat{\theta})}\right)$$

is often used as the statistic measuring evidence in favor of linkage, where $(\hat{\cdot})$ denotes the maximum likelihood estimate of the associated parameter(s) and μ denotes the mean common across genotype groups (Lander and Botstein 1989). Critical values that adjust for multiplicities across genome locations can be obtained theoretically (Dupuis and Siegmund 1999) or via permutations (Churchill and Doerge 1994).

The specific TB approach considered here assumes a Gaussian density for f_{obs} with critical values determined by the formulas given in Dupuis and Siegmund (1999). We consider the output from this approach at markers and refer to this as a TB marker regression (TB-MR) approach. The restriction to consider output only at markers is done to facilitate comparisons with MB methods, discussed below. For TB-MR, the genome wide type I error rate per transcript is controlled at 5% (Dupuis and Siegmund 1999).

0.2.3 Marker Based Approaches

To identify transcripts significantly linked to genomic locations, instead of testing each transcript for significant linkage across markers, one could test at each marker for significant linkage across transcripts. This amounts to identifying DE transcripts at each marker, with groups determined by marker genotypes. The MB approach refers generally to the repeated application, at each marker, of any method for identifying DE transcripts. In this setting, a number of approaches could be used. We here consider four.

The first is an empirical Bayes approach, *EBarrays*, with the log-Normal Normal model (LNN) described in detail in Kendziorski *et al.* (2003; 2004). This approach calculates the posterior probability of differential expression for every transcript. Thresholds can be chosen to control the expected posterior FDR across transcripts. For example, by specifying the threshold to be the smallest posterior probability such that the average posterior probability of all transcripts exceeding the threshold is larger than $1 - \alpha$, the posterior expected FDR is controlled at $\alpha \cdot 100\%$ (Newton *et al*. 2004). This marker-based empirical Bayes approach will be referred to as MB-EB. As in TB-MR, the LNN model assumes a Gaussian density for f_{obs}.

The second marker-based approach consists of obtaining p-values from a Student t-test followed by p-value adjustment; and the last two approaches consider moderated t-statistics followed by p-value adjustment. The details of the moderated statistic construction are given in Smyth et al. (2004) and Tusher et al. (2003), respectively. Adjustment for these last three methods is done using q-values to control the overall false discovery rate (FDR). In particular, to control the FDR at α, transcripts with q-values $<= \alpha$ are considered significant (Storey and Tibshirani 2003). MB-Q, MB-LIMMA, and MB-SAM will denote the three marker-based approaches, respectively.

0.2.4 Other Approaches

Although the TB and MB approaches are in many ways fundamentally different, they share an important flaw. Separate tests are conducted for each transcript-marker pair, and each measures evidence that the transcript maps to that marker relative to evidence that it maps nowhere. Since a transcript can map to any of many marker locations, the evidence that a transcript maps to a particular marker should not be judged relative only to the possibility that it maps nowhere, but rather relative to the possibility that it maps nowhere *or* to some other marker. This idea motivates the mixture over markers (MOM) model (Kendziorski *et al.* 2004). Briefly, MOM assumes a transcript t maps nowhere with probability p_0 or to marker m with probability p_m where $p_0 + \sum_{m=1}^{M} p_m = 1$ and M denotes the total number of markers. The marginal distribution of the data $\mathbf{y}_t$ is then given by

$$p_0 f_0(\mathbf{y}_t) + \sum_{m=1}^{M} p_m f_m(\mathbf{y}_t) \tag{0.1}$$

where f_m describes the distribution of data if transcript t maps to marker m (f_0 describes the data for non-mapping transcripts). The component densities are predictive distributions that can be derived under different parametric assumptions. For comparison, we take Gaussian observation components for the log measurements with Normal priors on the latent expression levels.

0.3 Evaluation of ETL Mapping Methods

The methods discussed above were evaluated using simulated data and data from two case studies. The simulations are in no way designed to capture the many complexities of ETL mapping data. Nevertheless, they do provide some insight into operating characteristics of each of the approaches. The first case study concerns an experiment in yeast and the second a study of diabetes in mouse.

0.3.1 Simulation

Recall that for a backcross population, a subject has one of two genotypes (AA or Aa) at each marker locus. For an F_2, three genotypes are possible (AA, Aa, or aa) and, as a result, a given transcript may be equivalently expressed (EE) or may be in any one of 4 DE patterns ($AA|Aa, aa$; $AA, Aa|aa$; $AA, aa|Aa$; $AA|Aa|aa$). Here $|$ denotes inequality among the latent genotype group means. We performed a simulation of an F_2 population in which pattern membership was determined by a multinomial where the expected proportion of transcripts in each DE pattern was specified at 3%, 3%, 1% and 3%, respectively (1% is used for the pattern that is least biologically plausible).

Care was taken to protect against biasing the results in favor of any of the methods considered. The details are given in Kendziorski *et al.* (2004). In short, a major difference among methods lies in the estimation of transcript variance ${\sigma_t}^2$. To set the variance for a simulated transcript t, we used the posterior mean of ${\sigma_t}^2$, given by $\frac{\sum_{k=1}^{n}(y_{t,k}-\bar{y}_{t,\cdot})^2+\nu_0{\sigma_0}^2}{\nu_0+n-2}$ (derived assuming the transcript specific variance is distributed as scaled inverse chi-square: ${\sigma_t}^2 \sim \text{Inv}\chi^2\left(\nu_0, {\sigma_0}^2\right)$). As $\nu_0 \to 0$, the posterior mean approaches $\frac{(n-1)s^2}{n-2} \approx s^2$, the transcript specific sample variance, which is the naive estimate of ${\sigma_t}^2$ for an EE transcript under TB-MR assumptions. Data simulated with small ν_0 is therefore consistent with assumptions made in TB-MR. As $\nu_0 \to \infty$, the posterior mean approaches a constant value ${\sigma_0}^2$, which is assumed in MB-EB (note that this assumption implies a constant coefficient of variation on the raw gene expression scale). By varying ν_0, operating characteristics could be evaluated without biasing the results in favor of one method. Data simulated by this empirical method had marginal distributions that were virtually indistinguishable from the observed data.

We consider a single ETL simulation with 100 animals and 2 chromosomes. Marker genotype data was obtained from chromosomes 2 and 3 of the F_2 data described

in the next section. Chromosome 2 (3) contained 17 (6) markers with an average intermarker distance of 7.6 (17.7) cM. An ETL at marker 5 on chromosome 2 was simulated; no ETL was simulated on chromosome 3. Seven sets of simulations were obtained for ν_0 between 5^{-5} and 5^5 (ν_0 for the actual F_2 data was estimated near 5). For each value of ν_0, 20 simulated data sets were generated. At each fixed ν_0, the profile marginal MLE was obtained for ${\sigma_0}^2$.

FDR gives the proportion of transcripts identified incorrectly as mapping to chromosome 2; i.e. they were EE or they were DE but mapped outside the region flanking the true ETL. Table 1 reports the operating characteristics. FDR is well above the target level of 0.05 for most methods and most values of ν_0. MOM is the only approaches capable of FDR control in this simple simulation setting. Power measures the ability to identify the DE transcripts exactly at marker 5 or either of the flanking markers which are 16.5 and 5.8 cM away, respectively. There is little variation in power across ν_0. MB-Q is the most powerful method, followed by TB-MR, MB-EB, and MOM. The difference in power between MOM and the others is statistically significant, but perhaps not *practically* significant as power is still near 80%.

As shown in Table 1, the results from MB-Q, MB-LIMMA and MB-SAM were very similar, most likely because the relatively large sample size (100 animals) yields statistics in MB-LIMMA and MB-SAM that have been "moderated" only slightly. A similar result was reported in Smyth *et al.* (2004), where an experiment with 16 animals was considered. For this reason, only results for MB-Q will be discussed hereinafter.

0.3.2 Case Studies

To further compare these approaches, we consider ETL mapping data from the yeast experiment described in Brem *et al.* (2002). It is structured as a backcross between a standard laboratory strain (BY) and a wild isolate from a California vineyard (RM). There are 6215 transcripts and 3312 markers. With only 40 segregants in the cross, recombinants are limited. We removed pairs of markers with fewer than 10 recombinants in between leaving 88 markers.

Brem *et al.* (2002) identified 8 regions enriched for linkage across the genome. Many transcripts in these hot spot regions have been at least partly validated using independent experiments. As noted in the Introduction, these regions are of much interest as they may contain a master regulator responsible for the control of transcripts sharing common biological function. A statistical test for enrichment of common function can done via *GOHyperG* in Bioconductor (Bioconductor Core Team 2004)). *GOHyperG* uses data from Gene Ontology (GO), where transcripts are categorized at varying levels of biological detail (the three broadest levels are molecular function, cellular component, and biological process - there are many subcategories within each). For a given set of mapping transcripts and a given function, a hypergeometric calculation is performed to test for enrichment of that function across the transcripts.

Interpretation of resulting p-values is not straightforward due to the many dependent hypotheses tested. Furthermore, the hypergeometric calculation tends to result in small p-values when GO nodes with few transcripts are considered. For these reasons, it has been suggested that one only consider interesting small p-values obtained from a relatively large set of transcripts (> 10) (Gentleman, 2005). Applying this criterion to the results from Brem *et al.* (2002) gives 5 regions, shown in Table 2.

Table 3 shows information similar to Table 2, for the top 5 regions (5 regions with the largest number of mapping transcripts) identified by MOM, TB-MR, and MB-Q. We see that TB-MR identifies 3 of the 5 regions identified by Brem *et al.* (2002) on chromosomes 3, 12, and 14. The location identified by Brem *et al.* (2002) on chromosome 2 is missed by TB-MR; and the location identified by TB-MR on chromosome 9 is not found using any other method and shows little evidence for enrichment of common function. This is likely a false positive. Similar results are obtained from MB-Q, with 3 of the 5 regions identified, and one potentially spurious identification on chromosome 8.

The MOM model performs better: 4 of the 5 regions identified by Brem *et al.* (2002) (on chromosomes 2, 3, 12, and 14) are also identified by MOM. The one region identified by Brem *et al.* (2002) but not MOM is a second location on chromosome 3. There are not enough markers considered (using the selected 88) to distinguish between these two regions using MOM. In addition to improved hot spot localization, MOM is generally more sensitive than the other methods. We suspect that the increased number of identifications made by MOM are not false discoveries as the additional transcripts maintain evidence for enrichment of the common function.

It is insightful to check the results from these approaches when control of particular error rates is not used for hot spot identification. For example, instead of defining hot spots in terms of the number of mapping transcripts (which depends on particular thresholds to generate binary calls), one could consider average evidence (across transcripts) of mapping at each location (average LOD, average posterior probability, or the average of 1 - q-value). Given hot spots identified in this way, one can simply rank transcripts at each hot spot by LOD score, posterior probability, or 1-q-value and then consider the top N transcripts for some N. In terms of regions identified and tests for enrichment of common function, we found results similar to those shown in Table 3 for N of 50 and 100.

The ETL mapping approaches were also evaluated using data from a study of diabetes in mouse. For details on the experiment, see Kendziorski *et al.* (2004). Briefly, it is well known that the *ob* mutation in the C57BL/6J mouse background (B6-*ob/ob*) causes obesity, but only mild and transient diabetes (Coleman and Hummel, 1973), while the same mutation in the BTBR genetic background (BTBR-*ob/ob*) causes severe type 2 diabetes (Stoehr *et al.* 2000). To gain insight into the genetic basis of these differences, a (B6 x BTBR)F_2-cross was generated yielding 110 animals. Selective phenotyping (Jin *et al.* 2004) was employed to identify 60 F_2 *ob/ob* mice. For each of the 60 mice, liver tissue was isolated and 45,265 mRNA abundance traits were collected at 10 weeks of age using Affymetrix Gene Chips (MOE430A,B). The probe

level data was processed using Robust Multi-array Average (RMA) to give a single, normalized, background corrected summary score of expression for each transcript (Irizarry *et al.* 2003). Low abundance transcripts, defined as transcripts with average expression level below the tenth percentile, were removed leaving 40,738 traits. Genotypes for 145 markers were also obtained (over 90% of the animals provided genotype data at any given marker).

Each method was applied to identify ETL. Hot spot regions are shown in the left panel of Figure 2. The first marker, D2Mit241, is adjacent to D2Mit9, which has recently been identified as an obesity modifier locus (Stoehr *et al.* 2004). Two additional regions identified by 4 of the 5 methods (on chromosomes 4 and 10) are not yet known to be involved in diabetes although we note that the region identified on chromosome 4 has been implicated in other analyses done in the Attie lab. The two regions identified by MOM alone on chromosomes 5 and 8 have been identified by other groups in earlier studies: D5Mit1 is a location known to affect triglyceride levels (Colinayo *et al.* 2003) and D8Mit249 is the marker on our map closest to the "fat" gene which is known to affect both diabetes and obesity (Naggert *et al.* 1995). This provides some evidence for the MOM approach, but much more biological validation is required.

It is interesting to note that the agreement between FDR controlled and rank based inferences observed for the yeast study was not observed here. Figure 2 (right panel) gives results from the diabetes case study using the binary scores. As shown, there is much less agreement across methods when the binary scores are used. We expect there are conditions under which averaging evidence across transcripts is more advantageous than reducing to a binary score (and vice versa). This is currently an area under investigation.

0.4 Discussion

The field of QTL mapping was reignited in the 1980's by advances that allowed for the relatively easy identification of genetic markers and their genotypes. Today, with major developments in high throughput technologies, a similar advance has taken place that allows for measurement of thousands of phenotypes. The number and nature of these phenotypes are what distinguish QTL from ETL mapping. In fact, ETL mapping is exactly traditional QTL mapping, but with thousands of expression traits considered as phenotypes. The simplicity with which this difference can be stated perhaps obscures the resulting challenges posed for the statistical analysis of ETL data.

When faced with just about any statistical problem, it is often best to first consider methods that are currently available. This was done for ETL mapping. The earliest ETL papers applied traditional QTL mapping methods to each transcript in isolation. Doing so does not account for multiple tests across transcripts; and we found this to have a real impact on increased FDR even in very simplified simulation settings. For some labs, an inflated FDR is tolerable as many genes can be tested quickly for

certain properties and discarded if found to be false positives. However, for many labs, such tests are prohibitively expensive and more appropriate statistical methods are needed.

The most recent ETL studies have made attempts at adjusting for multiplicities across both markers and transcripts using a two stage approach (Chesler *et al.* 2005; Hubner *et al.* 2005). The first stage obtains a single p-value for each transcript that is adjusted for multiple tests across markers; stage two controls the FDR across transcripts by calculating q-values from these p-values. With this approach, mapping transcripts are identified, along with the single most likely location to which these transcripts map. Preliminary simulation results (not shown) show very low power if attempts are made to control the FDR at 5%. This is consistent with the results reported in Chesler *et al.* (2005), where an FDR cutoff of 25% is used so that 101 transcripts can be identified (out of $12,422$ total transcripts).

Our general conclusion is that a clever application of statistical methods developed in the context of QTL mapping and/or multiple testing is not sufficient to address the complexities of the ETL mapping problem. As a result, we continue to investigate MOM. The MOM approach was designed explicitly to address the ETL mapping question. Operating characteristics evaluated via simulations as well as results from case studies are encouraging. Another nice feature of the MOM framework is that it can be extended to account for interval and multiple ETL mapping. This work is underway.

In summary, much more work is required before the analysis of ETL data becomes routine. In practice, we suggest an investigator apply a number of tools and focus initially on genomic locations at which most methods agree (such as the 4 regions shown in the left panel of Figure 2), keeping in mind that assumptions across different methods are often very similar and therefore by no means are the results of different methods independent confirmations. Statisticians can contribute to the ETL mapping effort by method development, evaluation, and validation; and by carefully considering those genomic regions that *do not* agree across methods. Such regions can provide valuable insights so that specific conditions under which different methods work best can be identified. Advances in each area and communication between the two are required to maximize the amount of information that can be derived from ETL mapping studies.

0.5 References

1. Brem, R.B., G. Yvert, R. Clinton, and L. Kruglyak. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science* 296, 752-755.
2. Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. (1980). Construction of a genetic map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32, 314-331.
3. Broman, K.W. (2005).Mapping expression in randomized rodent genomes. *Nature Genetics* 37, 209-210.

4. Bystrykh, L., E. Weersing, B. Dontje, S. Sutton, M.T. Pletcher, T. Wiltshire, A.I. Su, E. Vellenga, J. Wang, K.F. Manly, L. Lu, E.J. Chesler, R. Alberts, R.C. Jansen, R.W. Williams, M.P. Cooke, and G. de Haan. (2005). Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nature Genetics* 37, 225-232.

5. Chesler, E.J. L. Lu, S. Shou, Y. Qu, J. Gu, J. Wang, H.C. Hsu, J.D. Mountz, N.E. Baldwin, M.A. Langston, D.W. Threadgill, K.F. Manly, and R.W. Williams. (2005). Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nature Genetics* 37, 233 - 242.

6. Churchill, G.A. and R.W. Doerge. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963-971.

7. Coleman, D.L. and K.P. Hummel. (1973). The influence of genetic background on the expression of the obest (Ob) gene in the mouse. *Diabetologia* **9**, 287-293.

8. Colinayo, V.V., J-H. Qiao, X. Wang, K. Krass, E. Schadt, A. J. Lusis, and T. A. Drake. (2003). Genetic loci for diet-induced atherosclerotic lesions and plasma lipids in mice *Mammalian Genome* 14(7), 464-471.

9. Cox, N.J. (2004). An expression of interest. *Nature* 12, 733-734.

10. Dupuis, J., and D. Siegmund. (1999). Statistical Methods for Mapping Quantitative Trait Loci From a Dense Set of Markers. *Genetics* 151, 373-386.

11. Gentleman, R. (2005). Using GO for Statistical Analysis. Bioconductor vignette, 2005.

12. Hubner, N., C.A. Wallace, H. Zimdahl, E. Petretto, H. Schulz1, F. Maciver, M. Mueller, O. Hummel, J. Monti, V. Zidek, A. Musilova, V. Kren, H. Causton, L. Game, G. Born, S. Schmidt, A. Muller, S.A. Cook, T.W. Kurtz, J. Whittaker, M. Pravenec, and T.J. Aitman. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature Genetics* 37, 243 - 253.

13. Irizarry, R.A., B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, and T.P. Speed. (2003). Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics* **4(2)**, 249-264.

14. Kendziorski, C.M., M.A. Newton, H. Lan, and M.N. Gould. (2003). On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statistics in Medicine* 22, 3899-3914.

15. Kendziorski, C., M. Chen, M. Yuan, H. Lan, and A.D. Attie. (2004). Statistical Methods for Expression Trait Loci (ETL) Mapping. *submitted*, invited revision.

16. Lan, H., J.P. Stoehr, S.T. Nadler, K.L. Schueler, B.S. Yandell, and A.D. Attie. (2003). Dimension reduction for mapping mRNA abundance as quantitative traits. *Genetics* 164, 1607-1614.

17. Lander, E. S., and D. Botstein. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185-199.

18. Morley, M., C.M. Molony, T.M. Weber, J.L. Devlin, K.G. Ewens, R.S. Spielman,

and V.G. Cheung. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature* 430, 743-747.

19. Naggert, J.K., L. D. Fricker, O. Varlamov, P. M. Nishina, Y. Rouille, D.F. Steiner, R.J. Carroll, B.J. Paigen, and E.H. Leiter. (1995). Hyperproinsulinaemia in obese fat/fat mice associated with a carboxypeptidase E mutation which reduces enzyme activity. *Nature Genetics* **10**, 135-142.
20. Newton, M.A., A. Noueiry, D. Sarkar, and P. Ahlquist. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* 5, 155-176.
21. Sax, K. (1923). The association of size differences with seed coat pattern and pigmentation in Phaseolus vulgaris. *Genetics* 8, 552-560.
22. Schadt, E., S. Monks, T.A. Drake, A.J. Lusis, N. Che, V. Collnayo, T.G. Ruff, S.B. Milligan, J.R. Lamb, G. Cavet, P.S. Linsley, M. Mao, R.B. Stoughton, and S.H. Friend. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297-302.
23. Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 3(1), Article 3.
24. Stoehr, J.P., S.T. Nadler, K.L. Schueler, M.E. Rabaglia, B.S. Yandell, S.A. Metz, and A.D. Attie. (2000). Genetic obesity unmasks nonlinear interactions between murine type 2 diabetes susceptibility loci. *Diabetes* **49**, 1946-1954.
25. Stoehr, J.P., J.E. Byers, S.M. Clee, H.Lan, I.V. Boronenkov, K.L. Schueler, B.S. Yandell, and A.D. Attie. (2004). Identification of Major Quantitative Trait Loci Controlling Body Weight Variation in ob/ob Mice. *Diabetes* **53**, 245-249.
26. Storey, J.D., and R. Tibshirani. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100(16), 9440-9445.
27. Yvert, G., R.B. Brem, J. Whittle, J.M. Akey, E. Foss, E.N. Smith, R. Mackelprang, and L. Kruglyak. (2003). Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genetics* 35(1), 57-64.

0.6 Figures and Tables

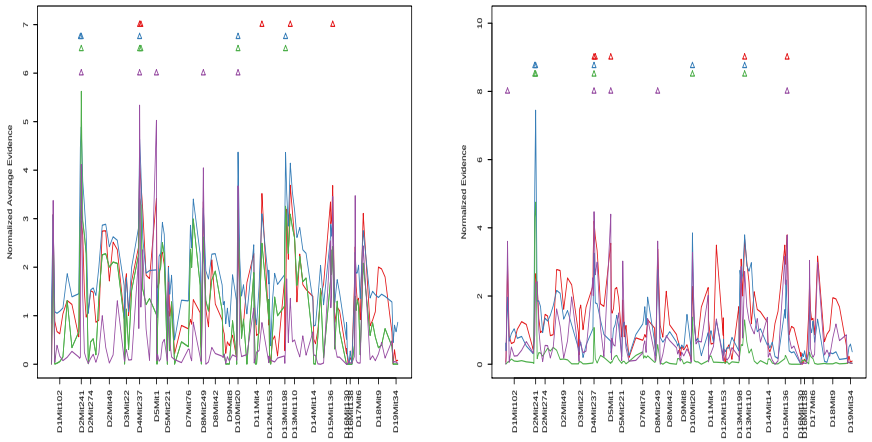

Figure 2: Evidence of linkage for each approach (LOD for TB-MR, posterior probability for MB-EB and MOM, and 1 - q-value for MB-Q). TB-MR, MB-EB, MOM, and MB-Q are colored by blue, red, purple and green, respectively. The left panel averages evidence of mapping over transcripts; the right panel gives normalized totals of mapping transcripts based on thresholding to control FDR. The 5 markers with the strongest evidence of mapping transcripts are indicated by triangles for each method.

OC	Method	ν_0						
		5^{-5}	5^{-3}	5^{-1}	5^{0}	5^{1}	5^{3}	5^{5}
FDR	TB-MR	0.286	0.286	0.293	0.285	0.286	0.28	0.301
	MB-EB	0.282	0.281	0.285	0.279	0.269	0.117	0.034
	MB-Q	0.24	0.246	0.246	0.24	0.245	0.23	0.226
	MB-LIMMA	0.238	0.236	0.232	0.237	0.235	0.237	0.229
	MB-SAM	0.233	0.238	0.235	0.232	0.238	0.236	0.221
	MOM	0.038	0.041	0.046	0.037	0.036	0.005	0.002
Power	TB-MR	0.884	0.886	0.887	0.886	0.889	0.919	0.868
	MB-EB	0.820	0.817	0.815	0.823	0.833	0.895	0.837
	MB-Q	0.911	0.912	0.913	0.912	0.917	0.949	0.918
	MB-LIMMA	0.900	0.910	0.909	0.900	0.914	0.935	0.899
	MB-SAM	0.897	0.908	0.906	0.898	0.913	0.933	0.899
	MOM	0.848	0.851	0.853	0.850	0.856	0.860	0.811

Table 1: Average operating characteristics (OCs) for TB-MR, MB-EB, MB-Q, MB-LIMMA, MB-SAM, and MOM. Averages are calculated over 20 data sets; standard errors were less than 0.005. OC definitions and details of the simulation are given in the text (see Section 0.3.1).

Chromosome (BP)	Number of Mapping Transcripts	Common Function	p-value
2(550)	18	Cell Separation	$\sim 10^{-7}$
3(90)	21	Leucine Biosynthesis	$\sim 10^{-7}$
3(190)	28	Mating	$\sim 10^{-10}$
12(670)	28	Fatty Acid Metabolism	$\sim 10^{-7}$
14(490)	94	Mitochondrial Induction	$\sim 10^{-6}$

Table 2: Results reproduced from Brem *et al.* (2002). Chromosomal locations, number of transcripts mapping to each region, biological function common to these transcripts, and p-values from GoHyperG are shown. BP gives the number of bases ($/1000$) from the 5' end of the chromosome.

Method	Chromosome (BP)	Number of Mapping Transcripts	Common Function	p-value
TB-MR	3(75)	29	Leucine Biosynthesis	$\sim 10^{-6}$
TB-MR	12(607)	21	Fatty Acid Metabolism	$\sim 10^{-7}$
TB-MR	14(502)	644	Mitochondrial Induction	$\sim 10^{-6}$
TB-MR	15(1)	27	Glucan Metabolism	> 0.2
TB-MR	9(99)	19	Iron Transport	0.03
MOM	2(602)	56	Cell Separation	$\sim 10^{-5}$
MOM	3(75)	56	Leucine Biosynthesis	$\sim 10^{-6}$
MOM	12(872)	55	Fatty Acid Metabolism	$\sim 10^{-8}$
MOM	14(502)	94	Mitochondrial Induction	$\sim 10^{-6}$
MOM	15(1)	288	Glucan Metabolism	$\sim 10^{-3}$
MB-Q	3(75)	31	Leucine Biosynthesis	$\sim 10^{-5}$
MB-Q	12(607)	36	Fatty Acid Metabolism	$\sim 10^{-7}$
MB-Q	14(502)	78	Mitochondrial Induction	$\sim 10^{-5}$
MB-Q	15(1)	29	Glucan Metabolism	10^{-1}
MB-Q	8(80)	81	Response to Pheromone	0.001

Table 3: Top 5 regions identified by TB-MR, TB-Q, and MOM. For each method and region, chromosomal locations, number of transcripts mapping to each region, biological function common to these transcripts, and p-values from GoHyperG are shown. BP gives the number of bases ($/1000$) from the 5' end of the chromosome. Note that the region identified by all methods on chromosome 15 is one of the 8 originally identified by Brem *et al.* (2002). It was excluded when constructing the list of 5 due to a relatively large p-value (0.02). It is difficult to judge whether or not this region is a false positive. Considering all methods point to this region, perhaps it is not.