Genome-wide Association Analysis in Diverse Inbred Mice:
Power and Population Structure

Phillip McClurg *,†† , Jeff Janes *,†† , Chunlei Wu *, David L. Delano *, John R.
Walker *, Serge Batalov *, Joseph S. Takahashi †,‡ , Kazuhiro Shimomura †,‡ , Akira
Kohsaka †,§ , Joseph Bass †,§,** , Tim Wiltshire *, and Andrew I. Su *

(*) Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121
(†) Department of Neurobiology and Physiology, (‡) Howard Hughes Medical Institute,
Northwestern University, Evanston, IL 60208
(§) Evanston Northwestern Healthcare Research Institute, Evanston, IL 60208
(**) Department of Medicine, Feinberg School of Medicine, Northwestern University, Evanston,
IL 60208

(††) These authors contributed equally to this work.

Microarray data from this article have been deposited in the GEO database at NCBI under the
accession number GSE5961.
Running head: Power and population structure

Keywords: association analysis, power, population structure, expression QTL

Corresponding author:

Andrew Su
10675 John Jay Hopkins Dr.
San Diego, CA 92121
858-812-1656 (voice)
858-812-1630 (fax)
asu@gnf.org
ABSTRACT

The discovery of quantitative trait loci (QTL) in model organisms has relied heavily on the ability to perform controlled breeding to generate genotypic and phenotypic diversity. Recently, we and others have demonstrated the use of an existing set of diverse inbred mice (referred to here as the “MDP”) as a QTL mapping population. The use of the MDP population has many advantages relative to traditional F2 mapping populations, including increased phenotypic diversity, a higher recombination frequency and the ability to collect genotype and phenotype data in community databases. However, these methods are complicated by population structure inherent in the MDP and the lack of an analytical framework to assess statistical power. To address these issues, we measured gene expression levels in hypothalamus across the MDP. We then mapped these phenotypes as quantitative traits with our association algorithm, resulting in a large set of expression QTLs. We utilized these eQTLs, and specifically cis-eQTLs, to develop a novel nonparametric method for association analysis in structured populations like the MDP. These eQTL data confirmed that the MDP is a suitable mapping population for QTL discovery, and that eQTL results can serve as a gold-standard for relative measures of statistical power.
INTRODUCTION

The use of modern genetics in model organisms relies heavily on the use of controlled breeding. Experimentally-derived crosses enable researchers to generate both genotypic and phenotypic diversity, and the resulting populations are analyzed to identify genomic loci which underlie disease and traits. Data from these experimental designs are analyzed using what are now classical methods of linkage analysis (Haley and Knott 1992; Lander and Botstein 1989). In rodent models over the past 15-20 years, these methods have been used to identify thousands of quantitative trait loci (QTL) across a wide range of phenotypes (Flint et al. 2005).

Recently, the increasing availability of high-throughput genotyping technologies has enabled the use of genome-wide association analyses for QTL discovery, and in some cases, in alternate mapping populations. For example, several research groups have investigated the use of a panel of diverse inbred strains of mice (Bogue and Grubb 2004) (collectively referred to here as the Mouse Diversity Panel (MDP)) for these QTL mapping studies (Cervino et al. 2005; Grupe et al. 2001; Liao et al. 2004; Pletcher et al. 2004). In contrast to mapping populations derived from controlled crosses, the strains of the MDP have been derived over the past century of semi-structured breeding and inbreeding (Beck et al. 2000). We recently introduced an algorithm for association analysis in the MDP based on the local inferred haplotype pattern, an approach we termed “haplotype association mapping” (HAM) (McClurg et al. 2006; Pletcher et al. 2004).

The MDP is an attractive choice as a mapping population for several reasons. Because these mouse strains were derived over the past century from crossing different mouse populations, the
MDP has a greater genetic and phenotypic diversity than is found in a typical F2 population derived from two parental strains. Since these mice are inbred, genotype data can be collected in community databases and applied to all mapping studies in the MDP. Finally, higher recombination rates and dense genotype maps result in more precisely defined QTL regions, facilitating the refinement of QTL to quantitative trait genes (QTG).

Nevertheless, there are also significant challenges of performing QTL studies in the MDP population. First, because we are limited to ~30-50 strains in the MDP for which we have dense genotype data, the question exists of whether there is sufficient statistical power to detect QTL (CHESLER et al. 2001; DARVASI 2001). Moreover, no analytical method for analyzing power has yet been developed. Second, the uncontrolled breeding process from which the MDP was derived can lead to spurious associations to background genetic structure if this population structure is not accounted for. Although the Collaborative Cross effort (CHURCHILL et al. 2004) will eventually address both these drawbacks experimentally, this mapping population is not yet available. In the mean time, QTL mapping in the MDP requires that these effects be accounted for analytically.

Ideally, computation of power and optimization of the association algorithm would utilize a set of positive controls where the QTG underlying a phenotypic trait has been unequivocally determined. However, in the case of genetic association analysis, particularly in the case of complex traits, the availability of these positive controls is severely limited (FLINT et al. 2005). Consequently, estimates of statistical power for QTL mapping studies typically rely on simulated genotype and phenotype data which are based on parametric assumptions. Although these
simulation studies produce useful estimates of statistical power, the parametric assumptions of normality are usually not satisfied in the typical real-world study. This limitation is especially clear when using association mapping in the MDP. Therefore, the degree to which these parametric estimates reflect the true statistical power in this population is unclear.

In this paper, instead of making parametric assumptions on simulated data, we used real experimental data and made assumptions on the identity of the true positives. The success of this approach clearly rested on the ability to identify a set of high-confidence true positives. Here, we utilized expression QTL (eQTL) data which maps gene expression measurements across the MDP as phenotypes for association analysis. cis-eQTLs are commonly considered to be a highly-enriched set of true positives with a low false positive rate (CHESLER et al. 2006). Although we did not know the identity of all genes which are truly cis-regulated (and hence could not calculate absolute measures of power), we used the presence of cis-eQTL as a relative measure of statistical power. We applied this approach to the development of an algorithm to account for the inherent population structure in the MDP. Furthermore, we assessed the ability of two-factor ANOVA models to improve the algorithm’s sensitivity.

Finally, since phenotype data measured in the MPD population is becoming increasingly available (BOGUE and GRUBB 2004), we also created a web site where users can submit phenotypes for analysis using the HAM algorithm. The URL is http://snpster.gnf.org.

RESULTS
**eQTL data:** Hypothalamus tissue was harvested from 32 strains of male mice from the MDP. Total RNA was isolated from frozen tissue and hybridized to a custom Affymetrix whole-genome mouse chip (GNF1M (Su et al. 2004)), and data were analyzed using the gcRMA algorithm (Wu et al. 2004). The 36182 probe sets represented on the chip were filtered to remove probe sets whose expression was either undetectable or invariant across strains. For each of the remaining 3725 probe sets, the expression pattern across strains was used as an input phenotype for the HAM algorithm.

Although each HAM analysis is performed completely independently of any knowledge of the gene’s position in the genome, a strong association is commonly observed between gene expression across the MDP and the haplotype pattern near the gene locus. This pattern is referred to as the “cis-eQTL band”, and is a robust pattern observed in many eQTL studies to date in a range of organisms and tissues (for example, Brem and Kruglyak 2005; Bystrykha et al. 2005; Chesler et al. 2005; Cheung et al. 2005; Hubner et al. 2005; Schadt et al. 2003; Stranger et al. 2005). Mechanistically, it is commonly hypothesized that cis-eQTLs are explained by a regulatory polymorphism that alters the ability of a transcription factor or enhancer to activate transcription, although allelic changes in mRNA stability and local DNA structure are also consistent with cis-regulation. Regardless of the specific mechanism underlying the observation, it is commonly assumed that there is a very low false-positive rate among cis-eQTLs (Chesler et al. 2006). Therefore, in this study we use cis-eQTLs as a large set of surrogate true positives, and conclude that methods which identify more cis-eQTLs are relatively more powerful.
We quantified the strength of the cis-eQTL signal by calculating an enrichment score, defined generally as the ratio of the eQTL peak density along the cis-eQTL band to the background genome-wide peak density. We defined cis-eQTLs to be any association within 500kb of the gene location (although evidence of cis-eQTL enrichment extends as far as five megabases, Supplementary Figure 1).

**Parametric versus nonparametric methods:** In this paper, we considered three variants of the original HAM method which differ in their approach to determining statistical significance. We first applied a **parametric HAM** method of measuring genetic association (McClurg et al. 2006). Briefly, this method used an inferred local haplotype based on a genotype calls over three contiguous SNPs. An F-test statistic was calculated based on the inferred haplotype grouping. Using the parametric version of HAM, significance of the F-test statistic was calculated from the theoretical background distribution, the F-distribution. This procedure was repeated for all three-SNP windows in the genotype data.

Using parametric HAM, we examined the top 10000 associations (nominal p < 1E-12) over all probe sets and SNP positions and found 19 associations within the +/- 500kb window defining the cis-eQTL band (Figure 1A). Compared to the size of the cis-eQTL band relative to the size of the mouse genome (1 megabase / ~2600 megabases), these results indicated a cis-eQTL enrichment factor of 4.94. Although the enrichment was modest, the results were statistically significant (p<0.01) based on eQTL analysis using 100 random permutations of the expression
data. None of these random simulations produced a cis-eQTL enrichment that exceeded the observed value, and the maximum enrichment observed in the background set was 2.08.

We next performed eQTL analysis using a nonparametric HAM version of our association method (McClurg et al. 2006). As before, an F-statistic was calculated based on the inferred haplotype pattern. Instead of calculating a p-value using a theoretical background F-distribution, we simulated the true background distribution using 1E6 bootstraps of the input phenotype. The eQTL map using nonparametric HAM is shown in Figure 1B. Of the top 10000 associations (10384 with ties, nominal p < 1E-6), 80 were within the cis-eQTL band, corresponding to a cis-eQTL enrichment factor of 20.03. Again, relative to 100 randomized eQTL simulations, this enrichment score was highly significant (p<0.01). To confirm that this trend is robust, we also examined cis-eQTL enrichment ratios using several other thresholds (Figure 2).

Based on its higher cis-eQTL enrichment factor, nonparametric HAM was more sensitive to detecting cis-eQTL associations, and therefore was considered a more powerful method. Nevertheless, it was apparent from Figure 1A and 1B that although the cis-eQTL band is statistically significant, it was not the most prominent signal in these eQTL maps. In both eQTL maps, a large number of horizontal bands were also clearly visible. Inspection of these horizontal bands revealed that this pattern reflects the case where the expression of one gene across strains was associated with the background population structure present in the MDP. Unlike F2 mapping populations in which parental haplotypes are shared with progeny in equal proportions, the MDP contains clusters of strains which are more related to each other than the other strains in the MDP. Clustering the global patterns of gene expression in hypothalamus
revealed a grouping of strains which closely matches the known ancestry of the MDP (Figure 3). In the eQTL context, nonspecific associations to this background population structure manifested themselves as horizontal bands.

**Accounting for population structure:** Here, we introduced a **weighted HAM** method to account for the population structure inherent in the MDP. To compute the association between a phenotype and an inferred haplotype pattern, the F-statistic was computed as usual. Next, significance was calculated using a weighted bootstrap procedure to simulate the null distribution. In nonparametric HAM, 1E6 bootstraps were performed in which a randomly chosen phenotype value was sampled from the vector of all phenotype values for each strain in the analysis. In the case of nonparametric HAM, the phenotype values from all strains were equally likely to be sampled as the replacement. Here, weighted HAM utilized the genome-wide genetic similarity to weight the sampling procedure. In this procedure, genetically similar strains were more likely to be selected as a replacement in the background distribution (details in Methods). This adjustment had the overall effect of increasing the distribution of association scores in the null distribution, thereby selectively decreasing the significance of nonspecific associations.

The effect of genome-wide genetic similarity correction can be adjusted using an empirical weighting exponent. An analysis with the weight exponent of zero exactly corresponded to an unweighted nonparametric HAM analysis (Figure 1B). Figure 4 displays the cis-eQTL enrichment as a function of the weight exponent. For all subsequent studies of the weighted HAM method, we used a weighting exponent equal to three (Figure 1C), chosen as the lowest
weighting exponent which visibly reduces nonspecific background association. Using the
weighted nonparametric HAM approach, we observed 190 cis-eQTLs among the top 10000
associations (10030 with ties, nominal p < 0.0004), corresponding to a cis-eQTL enrichment
factor of 49.25. Again, the improved cis-eQTL enrichment was evident over a wide range of
eQTL ranks (Figure 2), and was highly statistically significant relative to randomized eQTL
simulations (p<0.01).

**Trans-eQTL band enrichment:** The diagonal cis-eQTL band was the most striking pattern in
our hypothalamus eQTL data (Figure 1C) and was comprised of individual cis-eQTL peaks.
However, genetic associations to a gene’s expression pattern which do not map back to the
genomic position of the gene itself are called “trans-eQTL peaks”, and we also observed “trans-
eQTL bands” which form vertical lines in the eQTL plot. These patterns resulted from the case
where the expression patterns of multiple genes all associated to the haplotype pattern at a single
genetic locus. These patterns have also been observed in previous eQTL studies (BYSTRYKH et
al. 2005; CHESLER et al. 2005) and are presumed to reflect the situation where an upstream
transcriptional regulator at the QTL locus affects the expression of multiple downstream target
genes. As another distinctly non-random pattern which was likely to be enriched in true positive
associations, the occurrence of trans-eQTL bands was also used as a metric to evaluate relative
statistical power.

To determine the threshold of significance for the number of genes in a trans-eQTL band, we
performed 1000 bootstrap samples of the top 10000 association scores while randomizing
genomic position. This analysis indicated that trans-eQTL bands with greater than nine genes
were statistically significant (\(\alpha = 0.05\)). The results comparing the parametric, nonparametric, and weighted HAM methods are summarized in Table 1. This analysis showed that the weighted HAM analysis resulted in the most significant trans-eQTL bands among the three methods. Moreover, when comparing the size and significance of the largest trans-eQTL band from each method, the weighted HAM method also showed the best performance. Overall, analysis of trans-eQTL bands corroborated the findings of the cis-eQTL enrichment.

**Two-factor ANOVA model:** Since phenotype data for association analysis is commonly collected for both males and females separately, we next investigated the use of a two-factor ANOVA model to simultaneously model haplotype effects and sex effects. Although the application of a two-factor ANOVA model simply requires obtaining and phenotyping both male and female mice of each strain, the validation of the model using the concept of cis-eQTL enrichment required collection of gene expression data across strains for both sexes. Therefore, we chose a twelve-strain subset of the samples used in the hypothalamus study and performed an eQTL analysis. Clearly with only twelve strains, statistical power was drastically reduced. Over the top 500 associations, the cis-eQTL enrichment was reduced from 362 (corresponding to 27 cis-eQTLs) in the full 32 strain analysis to 26 (5 cis-eQTLs) in the twelve-strain subset. However, over 1000 randomized simulations, no computed cis-eQTL enrichment scores exceeded 16 (\(p<0.001\)), and enrichment threshold corresponding to \(\alpha = 0.05\) was only \(\sim 5\)-fold cis-eQTL enrichment. These simulations indicated that the cis-eQTL band in the twelve-strain subset was still highly statistically significant. (trans-eQTL analysis did not reveal a statistically significant signal in the twelve-strain analysis.)
We then obtained hypothalamus samples from female mice for each strain in the twelve-strain subset. Total RNA was isolated and gene expression was measured using our custom whole-genome Affymetrix array. Over the top 500 associations, the cis-eQTL enrichment in the female samples was 41.6 (8 cis-eQTLs), comparable to the score observed in the male samples. Next, we considered a two-factor ANOVA analysis over sex and haplotype. This model identified 27 of the top 500 eQTL peaks as cis-eQTLs, corresponding to a cis-eQTL enrichment ratio of 140.4. This comparison between the three eQTL runs clearly showed that the two-factor ANOVA model demonstrates much stronger enrichment of the cis-eQTL band (Figure 5). To confirm that this effect was not simply due to an increase in the number of animals phenotyped, we also performed a one-factor ANOVA analysis in which the sex factor was disregarded. This analysis showed cis-eQTL enrichment comparable to the individual sex-restricted one-factor analyses.

DISCUSSION

The prospect of using the MDP for genome-wide association mapping holds many potential advantages relative to more traditional mapping populations. In particular, the high recombination frequency between strains, coupled with dense genotyping datasets, leads to relatively precisely defined QTL with fewer candidate genes. Furthermore, because these strains are inbred, community efforts for genotype and phenotype data significantly decrease the cost of performing an association analysis for a phenotype of interest. Nevertheless, the two important liabilities of using the MDP as a mapping population are the lack of a framework to assess
statistical power and the background population structure. Here, we have introduced two approaches which address these issues.

Estimation of statistical power is an important aspect of characterizing any QTL mapping strategy. Several seminal studies of statistical power have been published which aid the interpretation of QTL results from F2 mapping studies (Darvasi and Soller 1997; Soller et al. 1976). However, these power estimates typically employ parametric assumptions on the structure of the data. In the context of F2 mapping populations, these statistical assumptions are usually reasonable due to the defined structure resulting from controlled breeding. However, the MDP is fundamentally different in a variety of ways. Most notably, this population is inbred with no heterozygous alleles, and the pedigree is unbalanced and largely uncertain.

Since the theoretical framework to compute absolute statistical power has not yet been worked out for the MDP, here we used cis-eQTL enrichment as a measure of relative power. Instead of assumptions on the structure of the data, this approach made assumptions on the identity of true positives. Here, we generated the first eQTL data set in the MDP population. The results were qualitatively the same as previously observed in mouse and in other organisms. In the eQTL context, cis-eQTLs can be considered a reliable set of true positives (Chesler et al. 2006), even after correcting for extensive multiple testing across genomic loci and gene expression probes (data not shown). Although it is not known how many genes are truly cis-regulated in the MDP (and hence we cannot calculate absolute power), it is likely that an algorithm that enriches for cis-eQTL is relatively more powerful. Furthermore, though the expression traits underlying cis-eQTLs are likely to result from alleles with very high genetic effect sizes, the algorithmic
optimizations and conclusions drawn from these studies will likely extrapolate to analysis of complex traits.

We first applied this power assessment metric to compare parametric and nonparametric calculation of association significance. Both methods utilized the same test statistic, but the former compared the computed test statistic against a theoretical background distribution whereas the latter simulated the background distribution using extensive bootstrapping. The nonparametric HAM method showed a substantially higher cis-eQTL enrichment, which indicated a higher statistical power relative to the parametric HAM. These results corroborated our previous findings which used F2-derived QTLs for a well-studied complex trait as a gold standard (McClurg et al. 2006). Moreover, these results underscored that phenotypic data in the MDP often do not follow typical parametric assumptions of normality.

We next applied the cis-eQTL enrichment metric to assess methods for correcting population structure effects inherent in the MDP. Previously, we accounted for these effects by altering the standard F-statistic using a weight factor reflecting genome-wide genetic similarity (McClurg et al. 2006). However, this method had an unknown effect on the test statistic’s null distribution. The challenges of association mapping in structured populations have also been previously addressed in larger population sizes by identifying unstructured subpopulations (Pritchard et al. 2000; Pritchard et al. 2000). Others have recently proposed an alternate method for accounting for population structure in the context of a mouse heterogeneous stock population using a “bootstrap posterior probability” (Valdar et al. 2006). However, because of the limited population size in the MDP, these previous efforts cannot be directly adapted to this mapping
population. Recently, another method was introduced in context of association mapping in an
Arabidopsis population similar to the MDP (TOOMAJIAN et al. 2006). As these authors alluded
to, their algorithm could be applicable to other structured mapping populations (including the
MDP).

Here, we have introduced an alternate method to generate a more appropriate null distribution of
the F-statistic in the MDP, thereby reducing the significance of p-values for nonspecific
associations. Since some strains in the MDP are clearly more related to each other than to the
rest of the strain set, our weighted bootstrap procedure utilized genome-wide genetic similarity
when selecting replacement phenotype values. This modification represented a different null
hypothesis being tested. The unweighted nonparametric method sought to reject the null
hypothesis that the association at a given locus was stronger than a random bootstrap of strain
labels. In contrast, the weighted HAM method tested the null hypothesis that the association at a
given locus was no stronger than the association to the global genome structure. To generate this
modified null distribution, strains sharing a high degree of genetic similarity were more likely to
be sampled when constructing a background phenotype vector. The use of a weighted
nonparametric HAM analysis clearly further improved the cis-eQTL enrichment metric.
Furthermore, this method resulted in a notable decrease in the horizontal bands in the eQTL map
which were indicative of nonspecific association to background genetic structure.

Finally, we have also applied the concept of cis-eQTL enrichment to evaluate the effect of a two-
factor ANOVA model in our HAM method. The incorporation of sex effects in regression
models for eQTL analysis has been previously described in a recombinant inbred mouse
population (WANG et al. 2006). Here, we showed that when gene expression data for both males and females were available, cis-eQTL enrichment was substantially improved relative to one-factor ANOVA models. We also performed a similar comparison of two-factor ANOVA in a second eQTL study across 23 strains and found similar results (data not shown). Although not all phenotypes will benefit from treating sex as a second factor in association analysis, the analysis of thousands of phenotypes in parallel indicated that there is an overall benefit. Incorporating the sex effect into the statistical model clearly improved the cis-eQTL enrichment relative to one-factor models. Nevertheless, the relative benefit versus the increased cost of additional phenotyping will likely vary depending on the specific experiment and phenotype being studied.

In summary, these results demonstrated that the use of the HAM algorithm in the MDP has sufficient statistical power to identify enrichment in both the cis-eQTL band and trans-eQTL bands. Moreover, we demonstrated that enrichment of these eQTL patterns is a suitable metric to calculate relative measures of power for genome-wide association mapping. We used this principle to design and optimize a novel method of accounting for population structure in the mapping population, and to evaluate a two-factor ANOVA variant of our HAM method. Finally, we created a web-based tool for performing HAM analyses using the weighted bootstrap method at http://snpster.gnf.org.

METHODS
**Sample Preparation:** All mice were maintained on a 12-h light/dark cycle at least 1 week before collecting hypothalami and individually housed with food and water available *ad libitum*. At 25 weeks of age, mice were sacrificed under light anesthesia at ZT 6 (ZT 0 defined as lights on) 2 h after food deprivation. Animals were sacrificed by cervical dislocation and hypothalami were taken by making two coronal cuts each just posterior to the optic chiasm and the pituitary stalk. A pair of sagittal cuts was made 1.5 mm from the midline. A final horizontal cut was made 1 mm dorsal to the floor of the hypothalamus. Tissues were snap-frozen in liquid nitrogen and stored at -80 °C for subsequent analysis.

**Gene expression data:** Hypothalamus samples were isolated from male mice of 32 strains (*n*=3 excepted as noted below): 129S1/SvImJ*, A/J*, AKR/J*, BALB/cByJ*, BTBR_T+tf/J*, BUB/BnJ*, C3H/HeJ, C57BL/6J*, C57BLKS/J, C57L/J, CAST/EiJ (*n*=2), CBA/J, CZECHII/EiJ, DBA/2J*, FVB/NJ, I/LnJ, JF1/Ms, MA/MyJ, MOLF/EiJ, MSM/Ms (*n*=2), NOD/LtJ (*n*=1), NZB/BlNJ*, NZW/LacJ*, PERA/EiJ, PL/J (*n*=2), RIHS/J*, SEA/GnJ, SJL/J, SM/J*, SPRET/EiJ, SWR/J, WSB/EiJ (*n*=2). RNA from male replicates were pooled prior to amplification and subsequently hybridized to a single chip per strain. Hypothalamus samples from female mice of 12 strains (indicated with “*” above) were also isolated (*n*=1). Gene expression analysis was performed according to standard procedures (Su et al. 2004). Briefly, RNA was isolated from frozen tissue using Trizol followed by cleanup with the RNeasy kit. RNA was amplified and labeled using the Affymetrix One-Cycle Target Labeling kit. Samples were hybridized to GNF1M whole-genome mouse arrays (Su et al. 2004), and data were processed using the gcRMA algorithm (Wu et al. 2004). Raw data were deposited in GEO (http://ncbi.nih.gov/geo) under series accession number GSE5961.
In this study, data were filtered to remove probe sets whose expression was either undetectable (maximum expression across strains less than 200) or invariant across strains (ratio of maximum expression to minimum expression across strains less than three). Results were qualitatively the same when no filtering for differential expression was performed. In addition, the results described in this manuscript were generated using the default CDF file. Although summarization algorithms are designed to be robust to single-probe outliers, the presence of SNPs in the probe sequence could theoretically lead to spurious detection of cis-eQTLs. An analysis performed after removing all probes overlapping a SNP in dbSNP from the CDF file resulted in qualitatively similar results (Supplementary Figure 2).

**Haplotype Association Mapping:** The original HAM method using inferred haplotypes has been previously described (McCLURG et al. 2006; PLETCHER et al. 2004). Briefly, the HAM approach inferred the haplotype structure from a set of contiguous genotype calls across the MDP, and used these inferred haplotype groups as the independent factor in a one-factor ANOVA. For each phenotype, the F-statistic was calculated to quantify the between-group variance relative to within-group variance. The background distribution used to estimate the significance of the test statistic was either computed parametrically (the F-distribution) or simulated nonparametrically using 1E6 bootstraps of the phenotype vector. The resulting p-value was then transformed using a –log10 transformation to produce an association score. As mentioned previously, we utilized a generalized family-wise error rate to control for multiple testing.
SNP genotype data were collected primarily from three sources: GNF (WILTSHIRE et al. 2003), Rosetta/Merck (CERVINO et al. 2005), and the Broad Institute (http://www.broad.mit.edu/~mjdaly/mousehapmap). All SNP locations were mapped to NCBIM33. In total, the genotype dataset contained allele calls for approximately 157,000 SNPs across 46 strains in the MDP, resulting in a median inter-SNP distance of 6.02kb across autosomes.

**Weighted Bootstrap:** We first calculated the genetic similarity matrix, as defined previously (MCCLURG et al. 2006). Briefly, this matrix calculated the pairwise similarity between two strains as the number of genotype calls in common divided by the total number of genotypes called in both strains. This matrix was “weighted” by raising all ratios to the exponent of the weight factor. A weight exponent of zero corresponded to a genetic similarity matrix of all ones.

As described above in the discussion of HAM, the significance of the calculated test statistic was estimated non-parametrically by simulating the null using bootstrapped phenotype values. A bootstrapped phenotype vector was created by choosing a replacement phenotype value for each strain in the phenotype vector. In the unweighted case, all strains were equally likely to be selected as a replacement for a given strain. In our weighted bootstrap, the probability of choosing the phenotype of a particular strain A as a replacement for strain B was proportional to the value in the genetic similarity matrix. When the weight exponent (described above) was set to zero, this corresponded to the unweighted case (since all values in the matrix are 1). As the weight exponent was increased, the likelihood also was increased of choosing a substitution strain in the null distribution that was genetically (and hence, phenotypically) similar. The net
effect was that the null distribution was increased for strong associations which are due to population structure, thereby selectively decreasing their significance and association score. This weighted bootstrap had a lesser negative effect on association scores which are not due to population structure.

**Cis-eQTL enrichment calculation:** The cis-eQTL enrichment score was calculated as the ratio of peak density in the cis-eQTL band relative to the overall peak density. We first applied a prefilter to remove multiple adjacent associations that reflect the same strain distribution pattern. First, the top genome-wide association was identified, and all association scores within 500kb were removed. The second-highest association score was similarly identified and isolated, and this process was iteratively applied for all peaks. After prefiltering, we had at most one representative for each 1 Mb interval for each probe set.

Cis-eQTLs were defined as associations between a gene expression vector and a genomic locus within 500 kb of the gene’s genomic location. At a given threshold, the percentage of cis-eQTL of all eQTL peaks was calculated. Since the defined cis-region was approximately 1/2600 of the overall genome size (~ 2600 megabases), the cis-eQTL enrichment was the percentage of cis-eQTLs multiplied by 2600.

**Trans-eQTL band calculation:** Trans-eQTL bands were detected by tabulating the number of genes whose expression was associated with a single genetic locus. The significance of each trans-eQTL band was calculated by comparing to a background distribution of trans-eQTL sizes.
This background distribution was generated by creating 1000 random bootstrap samples of the eQTL matrix.

**Two-factor ANOVA model:** The two-factor ANOVA model incorporated two main effects (haplotype and sex) and an interaction effect. For a genomic locus with three haplotype groups, the full model $F$ was:

\[
Y_{ijk} = \mu + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \epsilon_{ijk}
\]

The $X_{ijk}$ were indicator variables with values dependent on sex and haplotype. The error terms were independent and normally distributed. To test the significance of the haplotype effect, the full model was compared to a reduced haplotype model $R_{hap}$ in which the haplotype main effect terms were removed. A general linear test statistic was used to compare the variance explained by the full and reduced model.

\[
F^*_{hap} = \frac{SSE(R_{hap}) - SSE(F)}{df_{R_{hap}} - df_F} \div \frac{SSE(F)}{df_F}
\]

The $df$ terms were the degrees of freedom of the two models involved. Finally the significance of this statistic was assessed non-parametrically using our weighted bootstrap procedure to create a background distribution of $1E6$ random statistics. As in the one-factor case, the resulting p-value was transformed via $-\log_{10}$ to give an association score.

**ACKNOWLEDGEMENTS**
The authors thank Serge Batalov for assistance with constructing parsimony trees, and Michael Cooke and Richard Glynne for helpful comments and suggestions. Research was supported in part by NIH award U01 MH61915 (J.S.T.) and RO3 DK62056 (J.B.). J.S.T. is an Investigator in the Howard Hughes Medical Institute.

LITERATURE CITED


CERVINO, A. C., G. LI, S. EDWARDS, J. ZHU, C. LAURIE et al., 2005 Integrating QTL and high-density SNP analyses in mice to identify Insig2 as a susceptibility gene for plasma cholesterol levels. Genomics 86: 505-517.


FIGURE LEGENDS

Figure 1. eQTL plots for hypothalamus. eQTL plots were generated using three haplotype association mapping (HAM) methods: (A) parametric analysis, (B) nonparametric analysis, and


(C) weighted nonparametric analysis. In all plots, the x-axis represents the genomic SNP axis and the y-axis represents the genomic probe set axis. Each spot represents an association between the expression of a gene and the strain distribution pattern at a SNP location. Alternating colors indicate chromosome boundaries on the x-axis. In each plot, the top 10000 eQTL associations are shown.

Figure 2. Comparison of parametric and nonparametric HAM methods. This chart displays the cis-eQTL enrichment as a function of eQTL rank for each of three different HAM variants.

Figure 3. Clustering of gene expression data. This clustering dendrogram displays the relationship of global gene expression patterns between strains. Coloring of the strain names reflects clusters derived from clustering of genotype data. The clear relationship between global gene expression patterns and genome-wide genetic similarity underscores the need to account for population structure in association analyses.

Figure 4. Weight exponent analysis. To optimize the choice of the weight exponent, we calculated cis-eQTL enrichment using a range of weight powers and eQTL ranks. We chose a weight exponent of three for all subsequent studies.

Figure 5. Two-factor association analysis. Strain distribution pattern and sex were treated as independent factors in a two-factor HAM analysis. An eQTL analysis was performed using hypothalamus over twelve strains. For comparison, the cis-eQTL enrichment ratios are shown for the corresponding one-factor analyses.
Supplementary Figure 1. Defining the cis-eQTL boundary. This histogram displays the distance between a gene and the closest cis-eQTL peak when using that gene’s expression as a phenotype for HAM analysis. Distance bins between +/- 30 megabases are shown. Although substantial enrichment is observed as far as +/- 5 megabases from the gene location, we have chosen a relatively stringent threshold of +/- 500 kilobases (blue bars) as the boundary for defining a cis-eQTL.

Supplementary Figure 2. Analysis after removing probes overlapping known SNPs. To ensure that the gene expression patterns that were underlying cis-eQTLs were not the product of SNPs in the microarray probe sequence, we constructed a new CDF file which removed these probes from the analysis. In total, 33701 probes (4.6% of all probes on the microarray) were removed. gcRMA was used to summarize the data, and all analyses described in this manuscript were repeated. This figure illustrates that the results were qualitatively the same after these potentially confounding probes were removed.
Figure 1
Figure 2

- Weighted nonparametric
- Nonparametric
- Parametric
Figure 3
Figure 4
Figure 5

The graph illustrates the relationship between eQTL rank and cis-eQTL enrichment for different scenarios:
- Two factor, male and female
- One factor, male and female
- One factor, male only
- One factor, female only
Supplementary Figure 1
Supplementary Figure 2
Table 1. Trans-eQTL bands.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number significant ($\alpha=0.05$)</th>
<th>Largest trans-eQTL band</th>
<th>Size</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>parametric</td>
<td>0</td>
<td>7</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>nonparametric</td>
<td>1</td>
<td>10</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>weighted</td>
<td>25</td>
<td>28</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
</tbody>
</table>