DOUBLE EQTL MAPPING METHOD TO IMPROVE IDENTIFICATION OF TRANS EQTLS AND CONSTRUCT INTERMEDIATE GENE NETWORKS

by

TRACY LASHONDA SPROLDEN

(Under the Direction of Paul Schliekelman)

ABSTRACT

A double eQTL mapping method for identifying *trans* eQTLs that reduces multiple testing and increases the statistical power of eQTL mapping is described. The first round of the method involves eQTL mapping for each gene using only the peak SNPs located on the same chromosome as the gene. All of the peak eQTLs found in the first round of mapping were *cis* eQTLs. Next the genes are clustered into similar expression groups and a list of combined eQTLs is created for the group. A final round of eQTL mapping is done for each gene using the list of combined eQTLs for the group. The eQTLs that are found as a result of the second eQTL mapping were *trans* eQTLs. The genes and *trans* eQTLs identified from the double eQTL mapping method were then used to construct intermediate gene networks that are thought to link causative loci and traits or diseases.

INDEX WORDS: Single nucleotide polymorphism (SNP), Expression quantitative trait loci (eQTL), Microarray, *Cis* eQTL, *Trans* eQTL

DOUBLE EQTL MAPPING METHOD TO IMPROVE IDENTIFICATION OF TRANS EQTLS AND CONSTRUCT INTERMEDIATE GENE NETWORKS

by

TRACY LASHONDA SPROLDEN

BS, Valdosta State University, 2002

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF STATISTICS

ATHENS, GEORGIA

2013

DOUBLE EQTL MAPPING METHOD TO IMPROVE IDENTIFICATION OF TRANS EQTLS AND CONSTRUCT INTERMEDIATE GENE NETWORKS

by

TRACY LASHONDA SPROLDEN

Major Professor: Paul Schliekelman

Committee: Liang Liu
Cheolwoo Park

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

DEDICATION

I would like to dedicate the completion of this thesis to my two children, Daniel and Wanja. Throughout the difficult times, when it would have been easy to give up, they have been the source of my motivation to persevere.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Through expression quantitative trait loci (eQTL) mapping, markers that have a significant effect on gene expression can be identified. When eQTL mapping is performed, gene expression level is treated as a phenotype and markers are found that significantly affect the expression of each gene. Typically, eQTL mapping studies find predominantly *cis* eQTLs, genetic variants that affect gene expression that map close to the gene and often inside the gene (Yoav, G., *et al*. 2008, Hubner N., *et al*. 2005, and Doss S., *et al*. 2005). Genetic variants that significantly affect gene expression that are located further away or on different chromosomes from the gene are known as *trans* eQTLs. Typically, eQTL studies identify only small numbers of *trans* eQTLs. Generally it is assumed that the effects of *trans* eQTLs are pervasive and usually too small to be detected after the very large multiple testing correction (number of genes × number of markers) inherent to eQTL mapping. Though difficult to identify, *trans* eQTLs are important because they provide a closer look into downstream pathways that provide links between causative loci and traits or diseases.

Studies have been able to identify genes that are associated with causing traits or diseases, but the pathway between the causative loci, gene, and trait is still not clear (Sieberts S.K. and E.E. Schadt 2007). Several recent studies discuss the possibility of intermediate gene networks that link causative loci (eQTLs) and traits or diseases (Sun and Schliekelman 2011 and Fehrmann, R. S. N., *et al.* 2011). Most gene network research has been focused on identifying

single genes that impact traits or diseases (Sieberts, S.K. 2007). Recreating the structure of intermediate gene networks is difficult because the statistical link becomes harder to detect further away from the trait or disease (Fehrmann, R. S. N., *et al.* 2011). But in order to better understand complex traits or diseases, it will be essential to better understand the underlying causative gene networks (Sieberts, S.K. 2007).

In this study, a double eQTL mapping method is examined in order to reduce the number of test used for multiple comparisons and improve power, therefore focusing in on markers that are good candidates for being trans eQTLs. The first step of the eQTL mapping method involves mapping of each gene against the peak SNP(s) located on the same chromosome as the gene. The second step of the method requires the genes to be clustered into similar expression groups, create a list of peak eQTLs for the group, and finally complete eQTL mapping a second time for each gene using the compiled eQTLs for the group the gene belongs to. The eQTLs found in the second round of eQTL mapping will provide a list of *trans* eQTLs for each gene. Identification of *trans* eQTLs is essential to finding links between genes within intermediate gene networks in order to construct genetic pathways for traits or diseases, which is the focus of the next part of this thesis. The second part of this study utilizes the double eQTL mapping method to take a closer look at the structure of intermediate gene. The networks constructed for this research do not imply biological causal relationships among genes, but only implicate a statistical link between causative loci and gene networks.

CHAPTER 2

LITERATURE REVIEW

2.1 eQTL Mapping Studies

The purpose of eQTL mapping studies is to identify genomic locations where genotype significantly affects gene expression. eQTL mapping studies that involve yeast, flies, mice and humans have already been done (Schadt, E. E., *et al.* 2003, Brem, R. B., *et al.* 2002, and Morley, M. *et al.* 2004). Linkage and association mapping are the two types of eQTL mapping approaches. Linkage mapping seeks to identify genetic markers that predict traits and uses individuals with known relationships such as families or progeny of crosses between genetically divergent strains (Mackay, T. F, *et al*. 2009). The most predictive markers and causative loci for traits are located near each other and the probability of recombination (the formation of new combinations of genes) increases as genetic distance increases (Mackay, T. F, *et al*. 2009). Association mapping uses unrelated individuals from the same population and the locus variation has a direct effect on trait variation (Mackay, T. F, *et al*. 2009). In association mapping markers are tested individually and linkage mapping uses interval mapping to estimate the effect and map position of each eQTL (Mackay, T. F, *et al*. 2009). Regardless of the approach, eQTL mapping studies have shown large numbers of transcripts have genetic variablility (influenced by multiple eQTLs), expression levels of many transcripts are highly correlated, and *cis* eQTLs are more easily detected than *trans* eQTLs (Williams, R. B. *et al*. 2007 and Mackay, T. F, *et al*. 2009).

2.2 *Cis* and *Trans* eQTLs

Variants that regulate gene expression can be classified into two types, *cis* or *trans* eQTLs. Cis eQTLs have an effect on gene variation and are located in close proximity to the gene (Gibson, G. and B. Weir 2005). *Trans* eQTLs have an effect on gene variation, but are located farther away from the gene or on different chromosomes (Gibson, G. and B. Weir 2005). The majority of eQTL mapping studies have shown that cis eQTLs are more easily detected than *trans* eQTLs (Williams, R. B. *et al*. 2007 and Gilad, Y., *et al*. 2008). T*rans* eQTLs are difficult to identify because they tend to generally have smaller effects on gene expression, have higher false discovery rates than cis eQTLs, can be found anywhere in the genome, and increasing the stringency for eQTL detection decreases the ability to detect them (Gibson, G. and B. Weir 2005, Williams, R. B. *et al*. 2007, and Gilad, Y., et al. 2008).

CHAPTER 3

METHODS

3.1 eQTL Mapping Procedure

R programming language with the R/qtl package was used to carry out the eQTL mapping (http://www.r-project.org/). LOD (logarithm (base 10) of the odds) scores were calculated to determine linkage between each gene and SNP as described by Haley and Knott (1992). LOD scores compare the ratio of the probability of linkage to the probability of no linkage between each gene and SNP (Center 1995). The LOD scores were converted to χ^2 statistics by multiplying them by $2ln(10) \approx 4.6$, where χ^2 has two degrees of freedom, in order to determine the p-value of each LOD score (Center 1995). The p-values were then used to calculate the false-discovery rate (FDR) for each SNP (Benjamini and Hochberg 1995). The false-discovery rate was controlled at 0.1 in order to correct for multiple comparisons. Only SNPs with FDR < 0.1 were considered as eQTLs for each gene.

3.2 Grouping Genes

Hierarchical clustering (across genes) was used to separate genes into groups by expression similarity. A tree-cutting algorithm developed by Sun and Schliekelman (2011) along with the hclust function in R (http://www.r-project.org/) was used to group the genes. If more than 10% of the gene's expression data was missing, the gene was removed. The missing expression data for genes with less than 10% data missing was replaced with the mean

expression value for the gene. A dendrogram relating the genes is created using the expression data (clustering across genes), Euclidean distance, and complete linkage (Sun and Schliekelman 2011). A target group size is determined, four for this research but can be any number, and the genes are then divided into two groups, using R's cutree function, based on the genes location in the dendrogram (Sun and Schliekelman 2011). Repeated hierarchical clustering is applied along with cutree until groups of the target size or smaller are achieved (Sun and Schliekelman 2011). The resulting groups contain genes with correlated expression levels.

3.3 Data

The microarray data of Ghazalpour *et al*. (2006) was used for this research. A cross was conducted between inbred mouse strains C3H/HeJ and C57BL/6J. Liver gene expression data, genotype information at 1065 SNPs, and data on 22 physiological traits were collected using 135 female offspring from the cross. Ghazalpour *et al*. (2006) filtered the microarray data to the 3421 most varying and connected genes to use for their network analysis. The data were obtained from the authors' website (http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/MouseWeight/) and they are available to the public.

3.4 Identiifying *Trans* eQTLs

Various definitions exist for determining what makes an eQTL *cis* versus *trans*, all of which depend on the physical location of the gene in comparison to the eQTL peak. In order to determine a possible *cis* eQTL threshold, the distance from the middle of each gene locus to the the SNP with the highest LOD score (out of all SNPs on the same chromosome as the gene) was

calculated. There were 3279 genes with SNPs on the same chromosome as the gene and of these 869 genes had a SNP peak within 10 mb of the middle of the gene, 1331 genes had a SNP peak within 20 mb of the middle of the gene, 1700 genes had a SNP peak within 30 mb of the middle of the gene, and 2029 genes had a SNP peak within 40 mb of the middle of the gene. From analysis of the distribution of the distances from the middle of each gene to the peak SNP with the highest LOD score on the same chromosome as the gene, therefore only eQTLs located on different chromosomes from the gene as *trans* eQTLs for a gene.

A double eQTL mapping process was used to identify *trans* eQTLs for this research. In the first round of eQTL mapping each gene was mapped against only the peak SNP(s) on the same chromosome as the gene in order to reduce multiple testing and improve power. The peak eQTLs located on the same chromosome as the gene found from the first round of mapping will be referred to as *cis* eQTLs. Next, the genes were clustered into groups with similar expression levels and a combined list of eQTLs located on the same chromosome as the genes in each group was created. Finally, a second round of QTL mapping was done where each gene was mapped against the combined list of peak eQTLs (with the *cis* eQTLS found in the first round of mapping removed) for the group it belonged to. This greatly reduces multiple testing by focusing the *trans* eQTL mapping on markers that are good candidates for being *trans* eQTLs because they are *cis* eQTLs for other genes in the group.

CHAPTER 4

RESULTS

4.1 Comparison of eQTLs for Genes with Similar Expression Levels

The double eQTL mapping method requires the use and creation of similar gene expression groups with the assumption that the genes within the groups have common eQTLs. In order to verify this assumption, investigation of genes with similar expression levels was done to determine if those genes also had common eQTLs. The genes were clustered in groups containing a maximum of four genes per group, resulting in 1351 groups (333 groups with 1 gene, 331 groups with 2 genes, 343 groups with 3 genes, and 344 groups with 4 genes). The number of genes was reduced from 3421 to 3400 genes because 21 genes (with more than 10% of their expression data missing) were removed. eQTL mapping was done as described in Section 3.1 using the reduced set of 3400 genes and all 1065 SNPs.

The eQTLs for individual genes in the cluster groups were compared to determine the number of eQTLs shared by the genes in the group. The histograms in Figure 1 show the number of eQTLs shared between the genes in each group. One hundred random gene cluster groupings were also done to determine the validity of the shared eQTL results found from clustering the genes by expression level (Table 1). The number of shared eQTLs between groups with expression similarity was much greater than the number of shared eQTLs for randomly clustered groups. This result validates the assumption that genes within the correlated expression

groups tend to share common eQTLs and this supports the possible existence of genes and associated eQTLS working together in gene networks.

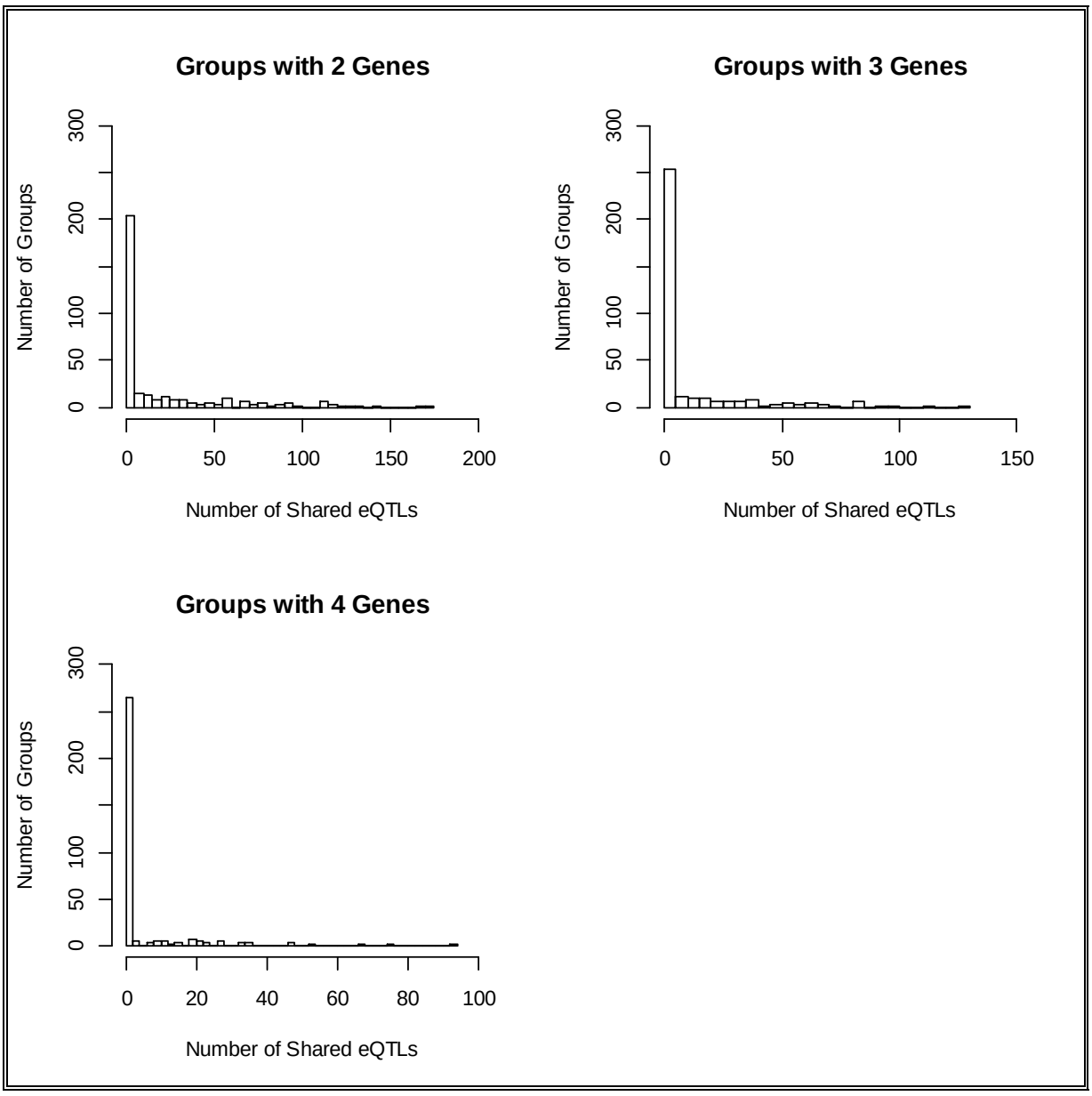

Figure 1: eQTLs Shared Between Genes in Cluster Groups

Table 1: Shared eQTLS for Expression Cluster Groups vs. Random Cluster Groups

Median 2 genes	Mean 2 genes	Max 2 genes	Median 3 genes	Mean 3 genes	Max 3 genes	Median 4 genes	Mean 4 genes	Max 4 genes	
0	0.27	31.86	0	0.01	3.16	0	0	0.15	Random 100 Average
0	3.89	95	0	1.91	60	0	1.49	75	Groups Using Expression

4.2 Traditional eQTL Mapping and Group eQTL Mapping

After it was determined that the genes in expression groups share common eQTLs, traditional eQTL mapping was compared to group eQTL mapping to determine if there was a difference in the number of genes identified with *trans* eQTLs by each method. Traditional eQTL mapping refers to mapping in which each gene was mapped against all 1065 SNPs and group eQTL mapping refers to mapping each gene against the combined eQTLs for similar expression groups. It will be necessary to identify genes with *trans* eQTLs in order to construct intermediate gene networks.

Traditional eQTL Mapping

First, eQTL mapping was done for each gene using all 1065 SNPs, which resulted in 1056 out of 3400 genes with associated eQTLs. The FDR significance threshold was set at 0.1, therefore only SNPs with FDR < 0.1 were considered as eQTLs. It is often the case that the eQTLs found from mapping include numerous eQTLs located in close proximity that are mapped to a gene, when in actuality there is only one (or two) peak eQTL(s) that are controlling gene regulation (Li and Burmeister 2005). In order to combat the problem of numerous eQTLs representing one peak, for each gene all eQTLs located on the same chromosome were

considered as one eQTL. For example, if a gene had ten eQTLs on chromosome three and five eQTLs on chromosome six, the gene was considered as having two eQTLs instead of 15 with one eQTL each from chromosomes three and five. Analysis of the 1056 genes with eQTLs showed 464 genes contained *cis* eQTLs (located on the same chromosome as the gene) and 922 genes contained *trans* eQTLs (eQTLs not located on the same chromosome as the gene). Of the 922 genes with *trans* eQTLs, 330 of those genes also had *cis* eQTLs. Figure 2 shows the distribution of independent (located on different chromosomes) *trans* eQTLs for the 922 genes.

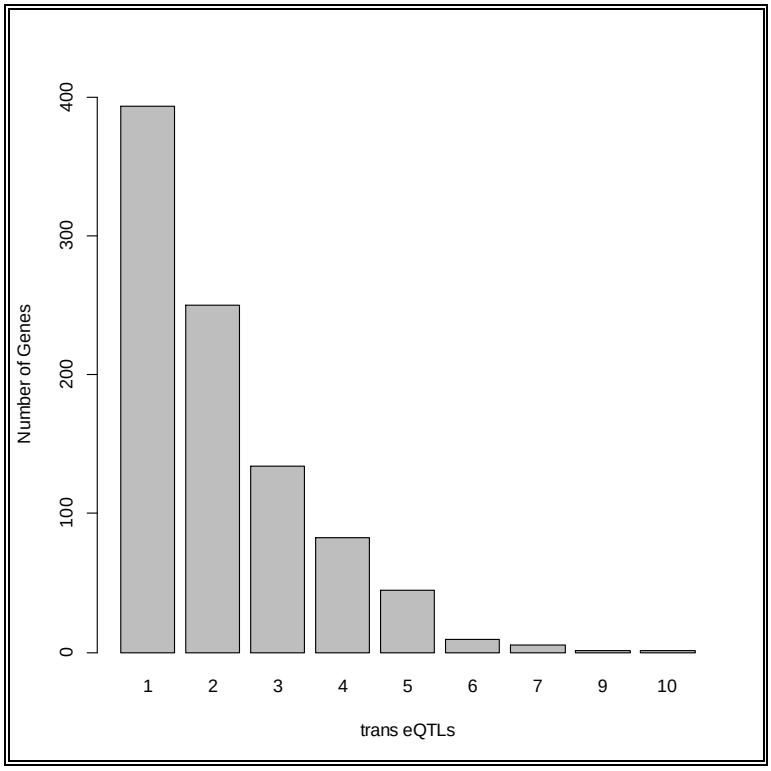

Figure 2: Distribution of *Trans* eQTLs from eQTL Mapping Using All SNPs

Group eQTL Mapping

Next the genes were clustered into similar expression groups in order to perform eQTL mapping using the combined eQTLs for similar expression groups. The genes were clustered into 1351 groups by expression level using the method explained in Section 4.1. The eQTL(s) for all genes in each group were then combined to create a list of all eQTLs for the group. Then eQTL mapping was performed for each gene using the combined list of eQTLs for the group with the originally mapped eQTLs removed for each gene, resulting in 723 genes with eQTLs. The 723 genes were found to belong to 389 groups (out of the 1351). Table 2 shows the distribution of "new relationships" found within the groups. A "new relationship" means that a new eQTL was mapped to a gene in a group after the group eQTL mapping was done that was not originally found in the traditional eQTL mapping. For example, suppose that three out of four genes in a group had eQTLs from the first mapping. Then the fourth gene and two of the other three genes in the group had eQTLs mapped to it from the second round of eQTL mapping, resulting in three "new relationships" created within that group.

Table 2: New Relationships Found from Group eQTL Mapping

Groups with 4 Genes (344 groups)					
New Relationships	0	1	2	3	4
Number of Groups	195	37	40	55	17
Groups with 3 Genes (343 groups)					
New Relationships	0	1	2	3	
Number of Groups	208	41	76	18	
Groups with 2 Genes (331 groups)					
New Relationships	0	1	2		
Number of Groups	226	84	21		

A list of the eQTL(s) found from the first eQTL mapping done using all SNP(s) and the eQTL(s) found from second QTL mapping done using the list of combined eQTLs for the group was created for each gene. The purpose of this was to create a complete list of eQTLs, from both QTL mappings, to analyze. There were 541 genes with *cis* eQTLs as the gene (originally 464 genes but some genes gained eQTLs from their group members) and 1309 genes with *trans* eQTLs. Of the 1309 genes with *trans* eQTLs, 404 of those genes also had *cis* eQTLs. Figure 3 shows the distribution of independent (located on different chromosomes) *trans* eQTLs for the 1309 genes.

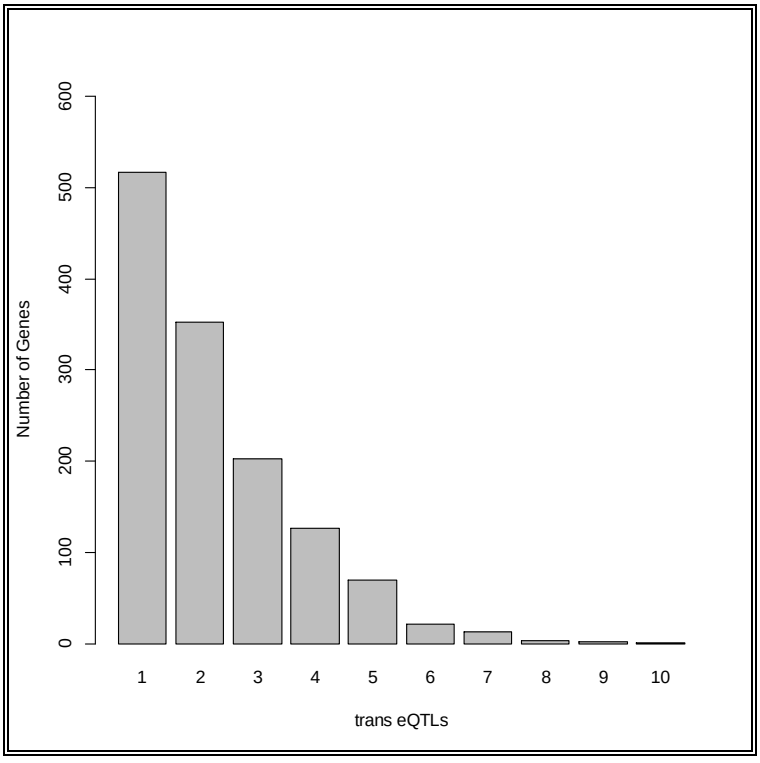

Figure 3: Distribution of *Trans* eQTLs from Group eQTL Mapping

Finally, traditional eQTL mapping was compared to group eQTL mapping to determine if there was an increase in the number of identified *trans* eQTLs. The traditional mapping process identified 922 genes with *trans* eQTLs, while the group eQTL mapping process identified 1309 genes with *trans* eQTLs. By using the group eQTL mapping process, 387 more genes with *trans* eQTLs were identified. Table 4 shows the distribution of *trans* eQTLS for both methods. These results support the assumption that group eQTL mapping increases the number of identifiable genes with *trans* eQTLs, thus resulting in an increase in the number of identifiable *trans* eQTLs for each gene.

Table 3: Distribution of *Trans* eQTLs for Traditional and Group eQTL Mapping

Traditional eQTL Mapping (922 Total Genes)										
trans eQTLs	1	2	3	4	5	6	7	8	9	10
Genes	394	250	134	83	45	9	5	0	1	1

Group eQTL Mapping (1309 Total Genes)										
trans eQTLs	1	2	3	4	5	6	7	8	9	10
Genes	517	352	203	127	70	21	13	3	2	1

4.3 Double eQTL Mapping Method

When group eQTL mapping was done using combined list of group eQTLs from mapping done using all SNP , there were a large number of *trans* eQTLs identified for each gene. Construction of the pathways and gene networks using the wide range of *trans* eQTLs discussed in Section 4.2 would be a daunting and difficult task. Therefore in an effort to reduce

multiple testing, improve power, and possibly increase the number of identified *trans* eQTLs, a double eQTL mapping method was created.

The first part of the double eQTL mapping process involved initial eQTL mapping for each gene using only SNPs located on the same chromosome as the gene. By using only the SNPs on the same chromosome as the gene, multiple testing is reduced from approximately 1000 tests per gene to about 100 tests per gene. There were 3258 genes with SNPs located on the same chromosome. (The 1065 SNPs were located on chromosomes 1-19 and the genes that did not have SNPs were located on chromosomes 0 and 20.) The LOD scores were then used to determine the peak (highest LOD score) SNP(s) located on the same chromosome as the gene. Then eQTL mapping was done using only peak SNP(s) on the same chromosome as the gene. The p-value for each peak SNP(s) along with the total number of SNPs located on the same chromosome as the gene were used for multiple testing to calculate FDR values for each SNP (Benjamini and Hochberg 1995). The FDR significance threshold was set at 0.1, therefore only SNPs with FDR < 0.1 were considered as eQTLs. Out of the 3258 genes with SNPs on the same chromosome as the gene, 536 genes were found to have significant *cis* eQTLs (peak eQTLs on the same chromosome as the gene).

Group eQTL mapping was shown to increase the identifiable *trans* eQTLs (Section 4.2) and constitutes the second part of the double eQTL mapping method. Clustering of the genes into 1351 groups by expression level was done using the method explained in Section 4.1. The *cis* eQTL(s) for all genes in each group were then combined to create a list of all eQTLs for the group. Next, eQTL mapping was done for each gene using the combined list of *cis* eQTLs for the group with the originally mapped peak *cis* eQTL(s).removed for each gene. For example, if there was a group of three genes and each gene contributed one *cis* eQTL to the list of combined

cis eQTLs for the group, the gene was mapped using only the remaining two eQTLs that it did not contribute to the group list. After the second eQTL mapping, 527 genes were found to have *trans* eQTLs. The 527 genes belonged to 281 out of the 1351 different groups.

4.4 Intermediate Gene Network Structure

In order to determine the structure of gene networks, the shared eQTLs in cluster groups was further analyzed. Of the 1351 gene networks (cluster groups) formed, there were 281 networks which contained genes that had *trans* eQTLs. Variation in the eQTL(s) located on the same chromosome as the gene lead to transcription changes in the gene network, which can then lead to variation in traits. Figure 4 shows some of the various ways the causative loci (*cis* eQTLs) and genes interact within networks. There were a total of 888 genes in the 281 networks (67 networks with two genes, 102 networks with three genes, and 112 networks with four genes). The results revealed insight into how eQTLs on the same chromosome as the gene and *trans* eQTLs work together in gene networks. There were 113 genes (found in 84 different gene networks) that did not contain eQTLs and therefore did not participate in the gene network. There were 343 out of 367 genes with *cis* eQTLs that provided eQTLS that were *trans* eQTLS for 527 genes. There were 41 networks with genes that "swapped" eQTLs (34 networks with two genes, 6 networks with three genes, and 1 network with 4 genes). Network 2 in Figure 4 provides an example of two genes that "swap" eQTLs. Table 4 provides a summary of the gene behavior observed within the 281 gene networks involved with the 527 identified *trans* eQTLs.

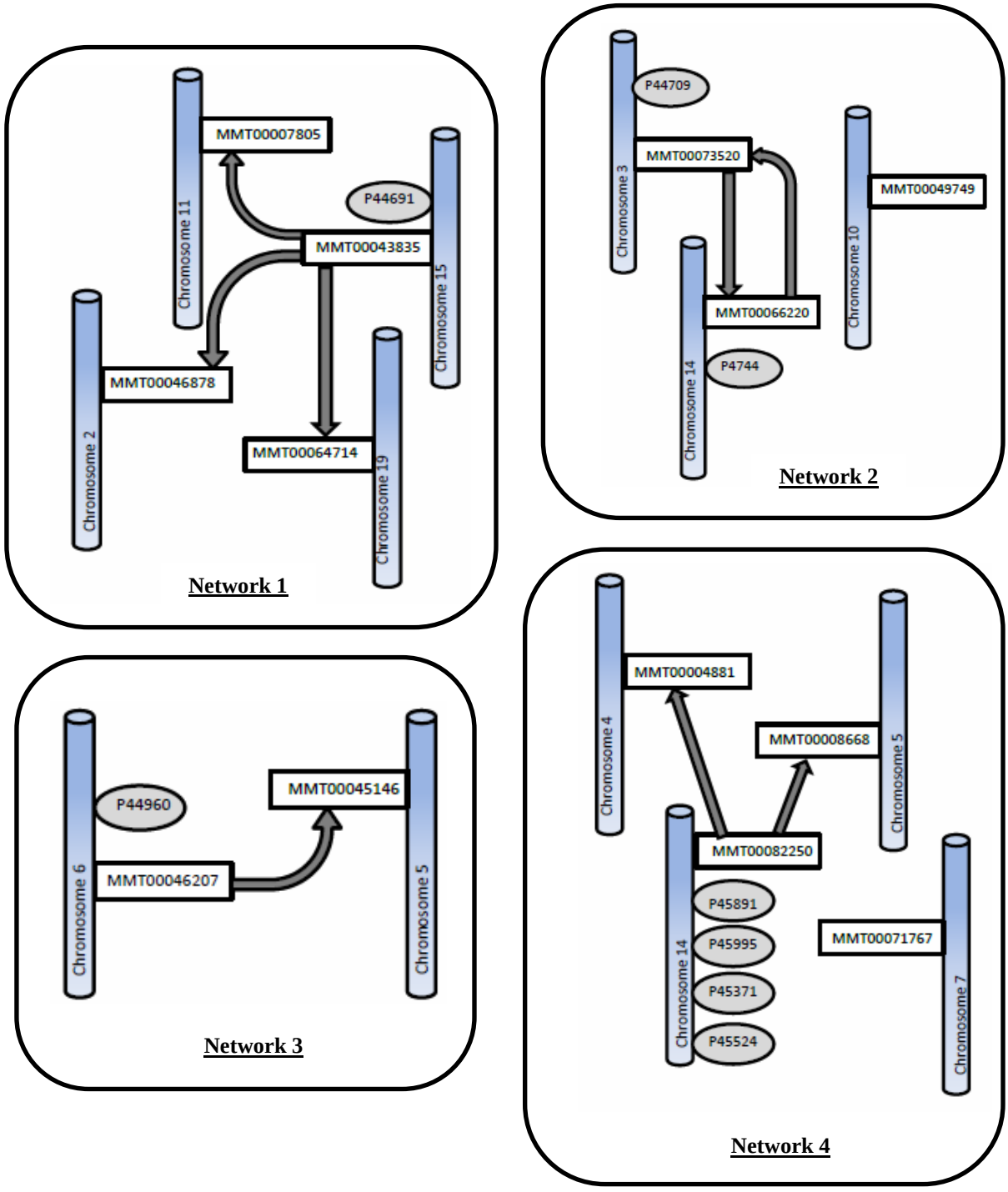

Figure 4: Intermediate Gene Networks. The ovals represent eQTLs on the same chromosome as the gene that lead to variation in the gene. Rectangles represent genes in a network with similar expression levels. The grey arrows between genes show how transcription is controlled by *trans* eQTLs.

Table 4: Summary of Gene Behavior within Gene Networks

	cis eQTLs	Gave eQTLS	Received *trans* eQTLs	No eQTLs	Swapped eQTLS
Number of Genes	367	343	527	113	90

CHAPTER 5

DISCUSSION

Here a double eQTL mapping method was introduced in an effort to reduce multiple testing, improve power, and increase detection of *trans* eQTLs for genes. In order to make sure the double eQTL mapping method was feasible, since it requires the genes to be separated into similar expression groups, it had to be determined if genes in similar expression groups share eQTLs. It was shown that number of shared eQTLs found for the similar expression groups was much greater than the number of shared eQTLs found when genes were randomly grouped (Table 1). The genes were then clustered into similar expression groups and eQTL mapping was performed for each gene using combined group eQTLS. The group eQTL mapping identified 1309 genes with *trans* eQTLS compared to the traditional mapping process which identified 922 genes with *trans* eQTLS, an increase in 387 genes with *trans* eQTLs. This result supports the creation of the double eQTL mapping method because it provides evidence of genes with similar expression levels sharing causative loci, which in turn makes it possible to identify more *trans* eQTLs using the similar expression groups.

The double eQTL mapping method introduced in this thesis, involves an initial round of traditional eQTL mapping (FDR < 0.1) for each gene using the peak SNP(s) on the same chromosome as the gene, creation of similar gene expression groups, and then a second round of eQTL mapping (FDR < 0.1) for each gene using a combined list of group peak eQTLs. The

double eQTL mapping method was shown to improve the ability to identify *trans* eQTLs for genes.

In order to reconstruct intermediate gene networks, that link causative loci to traits or diseases, *trans* eQTLs had to be identified. When all 1065 SNPs were used in the double eQTL mapping method, the resulting number of *trans* eQTLs was large making reconstruction of gene networks difficult. Instead of using all SNPs, only the peak SNP(s) located on the same chromosome as the gene were used in the initial eQTL mapping in an effort to control the number of identified *trans* eQTLs. Due to the fact the initial eQTL mapping only involved the peak eQTLs on the same chromosome, any eQTLs found in the second round of mapping (using the list of combined group eQTLs) will be *trans* eQTLs. The double mapping process was completed and 767 genes were found to have *trans* eQTLs. The eQTLs found in the first eQTL mapping where considered to be in close proximity to the genes and therefore were presumed to have a greater effect on variation in the genes. The *trans* eQTLs that resulted in the second round of mapping provided insight into the gene networks and identified genes that are affected downstream from causative loci. The constructed intermediate gene networks displayed various examples of how causative loci and genes work together (Figure 4 and Table 4). Some of the gene networks had only one gene contributing to the expression of other genes in the network, some had more than one gene contributing to the expression of other genes in the network, and still other networks had genes in the network that did not interact with the other genes in the group. Regardless of the underlying pathway within each gene network, the results support the existence of intermediate gene networks that connect causative loci and traits or diseases.

REFERENCES

Benjamini, Y. and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57: 289–300.

Brem, R. B., Yvert, G., Clinton, R. and Kruglyak, L., 2002 Genetic dissection of transcriptional regulation in budding yeast. Science 296: 752–755.

Center, C., 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature genetics 11: 241.

Doss, S., E. E. Schadt, T. A. Drake, and A. J. Lusis, 2005 Cis-acting expression quantitative trait loci in mice. Genome Research 15: 681-691.

Fehrmann, R. S. N., R. C. Jansen, J. H. Veldink, et al. 2011 *Trans*-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PloS Genet. 7(8): e1002197

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 in Genetics 21(11): 616-623.

Gilad, Y., S. A. Rifkin, and J. K. Pritchard, 2008 Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 24: 408-415.

Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315–324.

Hollingworth, P., D. Harold, R. Sims, A. Gerrish, J. C. Lambert, *et al.*, 2011 Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 43:429.

Hubner N., C.A. Wallace, H. Zimdahl, *et al*., 2005 Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. Nature Genetics 37: 243-253

Li J. and Burmeister M., 2005 Genetical genomics: combining genetics with gene expression Analysis. Hum. Mol. Genet. 14 (suppl 2): R163-R169.

Mackay, T. F., E. A. Stone, and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics 10(8): 565-577.

Morley, M., C. M. Molony, T. M. Weber, *et al.*, 2004 Genetic analysis of genome-wide variation in human gene expression. Nature 430: 743–747.

Schadt, E. E., S. A. Monks, T.A. Drake, *et al.*, 2003 Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297–302.

Sieberts, S.K. and E.E. Schadt, 2007 Moving toward a system genetics view of disease. Mamm Genome 18(6-7): 389–401.

Sun, G. and P. Schliekelman, 2011 A genetical genomics approach to genome scans increases power for QTL Mapping. Genetics 187: 939-953.

Williams, R. B., E. K. Chan, M. J. Cowley, and P.F. Little, 2007 The influence of genetic variation on gene expression. Genome Research 17(12): 1707-1716.

Yoav, G., S. A. Rifkin, and J.K. Pritchard, 2008 Revealing the architecture of gene regulation: the promise of eQTL studies. Trends in Genetics 24: 408-415.