Genome-level analysis of genetic regulation of liver gene expression networks

Daniel Gatti[1,¶], **Akira Maki**[1,¶], **Elissa J. Chesler**[2,¶], **Roumyana Kirova**[2], **Oksana Kosyk**[1], **Lu Lu**[3], **Kenneth F. Manly**, **Yanhua Qu**[3], **Robert W. Williams**[3], **Andy Perkins**[4], **Michael A. Langston**[4], **David W. Threadgill**[5,&], and **Ivan Rusyn**[1,&,*]

Daniel Gatti: dmgatti@email.unc.edu; Akira Maki: akiramaki.md@gmail.com; Elissa J. Chesler: cheslerej@ornl.gov; Roumyana Kirova: yordanovark@ornl.gov; Oksana Kosyk: kosyk@email.unc.edu; Lu Lu: lulu@nb.utmem.edu; Kenneth F. Manly: kmanly@tennessee.edu; Yanhua Qu: yqu@utmem.edu; Robert W. Williams: rwilliam@nb.utmem.edu; Andy Perkins: aperkins@cs.utk.edu; Michael A. Langston: langston@cs.utk.edu; David W. Threadgill: dwt@med.unc.edu

[1]Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

[2]Life Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831

[3]Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, TN 38163

[4]Department of Computer Science, University of Tennessee, Knoxville, TN 37996

[5]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Abstract

Liver is the primary site for metabolism of nutrients, drugs and chemical agents. While metabolic pathways are complex and tightly regulated, genetic variation among individuals, reflected in variation in gene expression levels, introduces complexity into research on liver disease. This study aimed to dissect genetic networks that control liver gene expression by combining large-scale quantitative mRNA expression analysis with genetic mapping in a reference population of BXD recombinant inbred mouse strains for which extensive SNP, haplotype and phenotypic data is publicly available. We profiled gene expression in livers of naive mice of both sexes from C57BL/6J, DBA/2J, B6D2F1, and 37 BXD strains using Agilent oligonucleotide microarrays. This data was used to map quantitative trait loci (QTLs) responsible for variation in expression of about 19,000 transcripts. We identified polymorphic cis- and trans-acting loci, including several loci that control expression of large numbers of genes in liver, by comparing the physical transcript position with the location of the controlling QTL. The data is available through a public web-based resource (www.genenetwork.org) that allows custom data mining, identification of co-regulated transcripts and correlated phenotypes, cross-tissue and –species comparisons, as well as testing of a broad array of hypotheses.

Keywords

eQTL; microarray; liver; mouse; genetical genomics

*Corresponding Author: Ivan Rusyn, M.D., Ph.D., Department of Environmental Sciences and Engineering, School of Public Health, University of North Carolina at Chapel Hill, CB #7431, Chapel Hill, NC 27599; Phone/Fax: 919-843-2596, iir@unc.edu.
¶Equally contributing first author
§Equally contributing senior author

Conflict of Interest
The authors declare they have no competing financial interest.

The maturation of gene expression technology has opened the door to the exploration of the genetics of gene expression [1-3]. Microarrays allow for the concurrent measurement of thousands of transcripts with the resultant genomic data being increasingly used to improve biological interpretation of data from mechanistic research. Phenotypic anchoring of observed phenotypes to gene expression changes has proven useful in uncovering the molecular mechanisms that lead to liver injury [4,5]. Such experiments connect the variation in transcript expression to phenotypes. However, they do not lead to detailed gene expression networks where the expression of one gene is found control the expression of another.

Recombinant Inbred (RI) mice are created by crossing two parental strains followed by sib-mating for over 20 generations [6]. Strains created this way have the advantage of being homozygous at almost every location along the genome. As such, each representative of an RI line will have limited phenotypic variation within that line, but the variation between lines is usually vast. RI panels are widely used to determine genotype-phenotype associations using QTL mapping techniques [7]. The relationships between phenotypes and genotypes are calculated using a likelihood ratio statistic (LRS), which is a measure of the probability that a given genetic marker explains the variation in the phenotype. When mRNA levels are used as the phenotype, regions of the genome with a high LRS are likely to contain genes that control the expression of the gene transcript being profiled; this process is referred to as expression Quantitative Trait Loci (eQTL) mapping [2,8].

The BXD panel of RI strains was created from the C57BL/6J and DBA/2J parental strains [9,10]. These two parental inbred strains are known to exhibit widely different phenotypes in response to a number of exposures. Thus, BXD mice have been a useful tool for elucidation of the genetic control of certain diseases. For example, the BXD lines have been used to study alcohol preference and tolerance [11], alcohol metabolism [12], responsiveness to aromatic hydrocarbons [13], N,N-diethylnitrosamine induced hepatocarcinogenesis [14] and diabetes and atherosclerosis [15,16]. Recently, BXD mice were used for eQTL studies that elucidated the genetics of gene expression in the brain [3] and hematopoietic stem cells [17].

The genetic basis and networks that control gene expression in the liver are not well delineated and improving our understanding of these pathways and controlling loci will advance knowledge of physiological and pathophysiological changes in this organ. Here, eQTL mapping was applied to data from genome-wide microarray profiling of liver gene expression in the naive state of the parental C57BL/6J and DBA/2J strains, the F1 cross and 37 BXD strains. Using this approach, potential key regulators of gene expression in the liver were identified. The comparison of the liver and brain transcriptome maps demonstrated tissue-specific differences in regulation of gene expression. Finally, we demonstrate how the data collected in this study may be used to infer genotype-phenotype correlations, generate testable hypothesis, and identify strains that may differ in responsiveness to xenobiotics based on the genetically-determined differences in expression of the key genes.

Experimental Procedures

Animals and Tissues

BXD1 through BXD42 mice are original RI strains available from the Jackson Laboratory (Bar Harbor, ME). BXD43 through BXD100 lines were generated using ninth or tenth generation Advanced Intercross Line (AIL) progenitors. AILs are generated by breeding two inbred parents (here is C57BL/6J and DBA/2J) and crossing their offspring so as to minimize inbreeding and maximize recombination events at each generation [10]. Mice were maintained at 20-24°C on a 14/10 hr light/dark cycle in a pathogen-free colony at the

University of Tennessee-Memphis. Animals were fed a 5% fat Agway Prolab 3000 rat and mouse chow and given tap water in glass bottles. Strain details are provided in Supplemental Table 1. Mice were raised to between 54 and 177 days (mean 70 days) of age. Liver tissues were collected following sacrifice by cervical dislocation. The whole liver was removed immediately, and placed in five volumes of RNAlater (Ambion, Austin, TX) at 4°C overnight, before removing from RNAlater and storing at -80°C until processing. All animal studies for this project were approved by Animal Care and Use Committee at the University of Tennessee-Memphis.

RNA Isolation

Total RNA was isolated from liver samples (~30 mg) using the RNeasy mini kit (Qiagen, Valencia, CA) as detailed by the manufacturer. RNA quality and quantity was determined spectrophotometrically from the absorbance at 260 nm and 280 nm. Aliquots of RNA samples were frozen at −80°C until microarray analysis. For each microarray, RNA was pooled from 2-3 mice of the same sex and strain.

Microarray Analysis

One μg of total RNA from 3 individual mouse liver samples of the same sex and strain was pooled, amplified and labeled with a fluorescent dye (Cy5) and a common reference - pooled mouse tissues (equal amounts of RNA from liver, kidney, lung, brain and spleen) mRNA isolated from the livers of 100 male C57BL6/J mice [see [18] for details] - were amplified and labeled with Cy3 using Agilent Technologies (Palo Alto, CA) Low RNA Input Linear Amplification labeling kit following the manufacturer's protocol. The quantity of the resulting fluorescently labeled cRNA was measured using a Nanodrop ND-100 spectrophometer (Nanodrop Technologies, Wilmington, DE) and its integrity assessed using an Agilent Bioanalyzer. Equal amounts of Cy3 and Cy5-labeled cRNA (750 ng) from the individual animals and from the pooled control, respectively, were hybridized to an Agilent Mouse Oligo Microarray (~21,000 features, catalogue# G4121) for 17 hours at 65°C. The hybridized microarrays were then washed and scanned using an Agilent G2565BA scanner. Data were extracted from the scanned image using Agilent Feature Extraction software version 6.1. A total of 122 arrays were run in 8 batches. The samples were semi-randomly distributed throughout the batches prior to microarray analysis in order to separate sexes and strains, and to minimize between- and within- batch bias. Technical and biological replicates were run both within each batch and between batches. The microarray data was deposited in the UNC Microarray Database and extracted using Log2 ratios of the mean red channel intensity over the mean green channel intensity. This was followed by LOWESS normalization to remove the intensity dependent dye bias [19]. Neither the genes nor the arrays were centered. Inter-batch normalization was carried out using a nested ANOVA mixed model with samples within each batch crossed with sex and strain.

QTL Analysis and WebQTL

QTL linkage mapping was carried out using the QTL Reaper software package (qtlreaper.sourceforge.net). One thousand permutations of the strain labels were performed to estimate the genome wide p-value [20]. Liver expression data was deposited in WebQTL (www.genenetwork.org) which is a web-based resource for exploring gene expression and phenotype interactions. WebQTL was used to produce interval maps for specific genes.

Transcriptome Map and Whole Genome QTL Clustering

The transcriptome map of the liver was produced using the R statistical package using the output data from QTL Reaper which consisted of the maximum Likelihood Ratio Statistic (LRS) value for each transcript on the microarray as well as a permutation derived p-value

(number of permutations = 1000). The p-value threshold was applied at a 25% False Discovery Rate [21]. The QTL data for all informative markers and 18,716 transcripts on the G4121A microarray was used to perform hierarchical clustering of the QTL data. Transcripts were clustered using each 2,325 marker vector of LRS values. The distance metric used for complete (maximum distance) hierarchical clustering was "1 – Pearson correlation" and colored by normalized LRS value.

Transcription Factor Analysis

Three web-based tools were used to search for possible transcription factor binding sites in candidate loci: the National Cancer Institute's Advanced Biomedical Computing Center promoter analysis tool (grid.abcc.ncifcrf.gov/promoters.php), oPossum version 1.3 [22] and PAINT [23]. In the first two cases, transcripts were divided based on increased expression correlating with the C57BL/6J of DBA/2J allele. The transcripts were submitted in 4 groups: 1) high expression with the C57BL/6J allele and an LRS of 30 or greater, 2) high expression with the C57BL/6J allele and an LRS of 40 or greater, 3) high expression with the DBA/2J allele and an LRS of 30 or greater and 4) high expression with the DBA/2J allele and an LRS of 40 or greater. With PAINT, all transcripts were submitted as one list, but a gene cluster file was also submitted that clustered the C57BL/6J high and DBA/2J high transcripts into separate clusters.

Quantitative Real Time PCR analysis

In order to test the hypothesis that Dicer1 might be the regulatory gene at the distal Chr12 locus which controls the 111 transcripts, female Dicer1 wild type and heterozygote mice [24] were obtained. No knockout is available because Dicer1 is embryonic lethal. 10 genes that are trans-regulated at this locus were selected (*Abhd1, Dhrsx, Hnf4g, Met, Neurog3, Olfr656, Pms1, Runx3, Rqcd1, Sel1h*). 9 genes were selected that are known to be expressed in the murine liver but do not have a significant or suggestive QTL on distal Chr 12 (*Abcb11, Abcc2, Abcd3, Cyp2e1, Cyp3a11, Cyp7a1, Cyp8b1, Ppar1, Rara*). RNA isolation was carried out using the Qiagen (Valencia, CA) RNeasy Mini Kit 50. RNA concentration and quality was checked using the Nanodrop (Wilmington, DE) ND-1000 spectrophotometer at the 260 and 280 nm wavelengths. 20 μg of RNA were used to produce cDNA using the Applied Biosystems Inc. (Foster City, CA) High Capacity cDNA Archive Kit. The Stratagene (La Jolla, CA) FullVelocity QYBR Green QPC Master Mix was used to perform the RTPCR and the plates were run on a Stratagene Mx3000P instrument.

Results and Discussion

Genetic control of gene expression in liver

Using web-based eQTL mapping tools and data collected in this study, regulatory loci controlling each liver transcript can be easily visualized in WebQTL (www.genenetwork.org). Figure 1 shows examples of the three types of expression control found in the liver. A transcript with a maximum locus of control near (+/- 20 Mb) the genomic location of the transcript itself is considered to be *cis*-regulated [25]. This implies some mechanism of control near the gene itself; perhaps a polymorphism in the promoter region. An example of a strong cis-regulated transcript is the cytochrome P450, family 3, subfamily a, polypeptide 13 (*Cyp3a13*) gene, located on distal chromosome 5 (Figure 1a). A transcript with a maximum locus of control far from the gene location is *trans*-regulated. This implies regulation by another gene like a transcription factor. ADP-ribosyltransferase 5 (*Art5*) gene, located in the middle of chromosome 7, is an example of a trans-regulated gene whose expression levels are determined by polymorphisms on Chr 3 (Figure 1b). Interestingly, many liver-expressed transcripts are regulated by multiple loci. The interleukin 21 receptor (*Il21r*) gene, located on distal chromosome 7, is a good example

(Figure 1c). *Il21r* has two significant QTL peaks - one on chromosome 7 proximal to the gene itself and one on chromosome 14.

To visualize the patterns of genetic control of gene expression on a genome-wide level, the 18716 annotated transcripts were clustered using the LRS vector for each transcript (Figure 2a). As expected, the majority of the transcripts in liver are independently regulated; however, several distinct patterns emerge. Specifically, there are a number of clusters of transcripts that all share a common maximal QTL as well as clusters that are co-regulated by a complex set of common loci. We refer to clusters regulated by only one strong QTL as "simple" QTL clusters and ones regulated by multiple loci as "complex" clusters.

Chromosome 8 contains a simple cluster of transcripts that all have a strong maximum QTL (mean LRS = 47.5, Figure 2b). Of the 27 transcripts in this cluster, 26 are located on chromosome 8 at the same location as the maximum QTL which indicates that this cluster contains predominantly *cis*-regulated genes or perhaps is due to a strain-specific difference in a regional transcriptional enhancer. Interestingly, the presence of one of the parental (C57BL/6J or DBA/2J) allele at this locus strongly affects expression of these genes (Figure 2b, yellow-red correlation plot).

Chromosome 12 also contains a simple cluster that consists of a set of transcripts with a maximal QTL (mean maximum LRS = 42.6) at the distal region of the chromosome (Figure 2c). However, only 5 of the 111 transcripts in this cluster are *cis*-regulated and hence located under the maximal QTL. This region in the genome is a *master*-regulator since the majority of the transcripts are *trans*-regulated by this locus. Again, the correlation between expression levels of these transcripts is strongly dependent on the type of the parental allele that is present at this locus implying a common regulator of all 111 transcripts.

To further understand the genetic basis for the difference in the proportion of cis- and trans-regulated transcripts in simple clusters, 31 simple clusters comprised of at least 80% cis-regulated transcripts were examined. It was hypothesized that the clustering of these cis-regulated QTLs might be due to higher gene or SNP density as opposed to biological pathways. To determine why a large number of cis-regulated transcripts might be located in very small genetic loci gene density at each region was considered versus adjacent up- and down-stream regions. For gene density, it was found that the QTL cluster regions had 79.5 ± 70.0 (mean $\pm$ SD) genes while the immediately adjacent up- and down-stream regions contained 45.0 ± 48.3 and 45.0 ± 44.2 genes, respectively. For SNP density, it was found that the QTL cluster regions had 4362 ± 3707 SNPs and the up- and down-stream regions had 2924 ± 2963 and 3318 ± 3957 SNPs, respectively. In this analysis, only those SNPs that differ between the C57BL/6J and DBA/2J strains in the Perlegen mouse SNP data set (mouse.perlegen.com/mouse/index.html) were considered. Thus, on the level of individual clusters, QTL clustering may be driven by a combination of gene and SNP density. Furthermore, the transcripts in these cis-regulated clusters do not appear to be enriched for any particular GO category. This indicates that the co-regulation of cis-controlled genes throughout the genome is not due to functional relatedness, but rather the high gene/SNP density in each region.

Lastly, a cluster of 43 transcripts (Figure 2d) that are controlled by a complex pattern of loci across multiple chromosomes (mean maximum LRS = 13.3) is shown. Not surprisingly, these genes are scattered around the genome. The pair-wise gene expression correlation matrix for these transcripts shows that mRNA levels for these genes are highly positively correlated regardless of the allele type at each QTL.

The proportion of cis-eQTLs that had higher expression when the C57BL/6J allele is present was assessed. Following the criterion set out in Pierce et al. [26], a cis-eQTL was defined as a

transcript that has a maximum QTL within +/- 5Mb of the transcript's location in the genome. At a genome wide p value of 0.05, 1,255 cis-eQTLs with 54.3% were found as having C57BL/6J high expression. At higher levels of statistical stringency 1,075 cis-eQTLs (p = 0.01) are found with 53.9% higher for the C57BL/6J allele and 867 cis-eQTLs (p = 0.001) with 53.3% higher for the C57BL/6J allele. The data presented by Doss et al [25] and the data from this study both show only a slight (54-56%) enrichment for C57BL/6J high cis-eQTLs when Agilent long oligo arrays, produced from the reference sequence of the C57BL/6J strain, are used. Consequently, much higher (72-75%) enrichment in the Pierce et al. [26] data set could be attributed to a low fidelity binding of short oligos used on Affymetrix arrays. These results show that long oligo array platforms are more suitable for eQTL analysis.

Mouse brain and liver transcriptomes show little overlap in genetic regulation of gene expression

Several recent reports identified a number of master-regulatory loci in other mouse tissues [3,17,26]. Here, the mouse brain (forebrain) and liver transcriptome maps (Figure 3) are compared to uncover the similarities and differences in genetic regulation of gene expression across tissues. Both brain and liver contain genes that are strongly *cis-* or *trans-*regulated at single loci, or are regulated by multiple loci (Figures 3a and b, respectively). In the brain transcriptome, three distinct master-regulator trans-bands are located on chromosomes 1 and 2 (Figure 3c). In the liver transcriptome, the strongest trans-band is located on distal chromosome 12 (Figure 3d), a locus that does not appear to be controlling expression of a large number of genes in the brain. While the liver and the brain both have a trans-band on chromosome 2 near 125 Mb, the two bands are not coincident. The liver trans-band lies at 119 MB and the brain at 135 Mb - a difference of 16 Mb.

Next, the QTL data for the liver and brain by selecting only those transcripts with a genome-wide p value 0.05 [26] was filtered. This analysis identified 743 transcripts that have significant QTLs in both tissues (Figure 3e). It is interesting to note that 209 transcripts are regulated by the same genomic region in both the liver and the brain (genes that fall of a diagonal in Figure 3e). Gene Ontology-based [27] analysis of the biological processes that are significantly over-represented (Fisher's exact test $p < 0.05$ and 25% FDR) among these genes identified a number of significant categories that are co-regulated in both tissues (Figure 3f). Furthermore, among the 743 transcripts, the proportion of cis-regulated transcripts in the liver was 0.59 and in the brain it was 0.55. The difference is due to slight differences in the transcript location between the two data sets. Of the 209 transcripts that have the same QTL in the liver and the brain about 95% are cis-regulated. Collectively, this comparison indicates that important tissue-specific patterns of genetic control of gene expression can be elucidated by this approach and potentially form the basis for comparative analyses between tissues.

Chromosome 12 contains a strong liver-specific master-regulatory locus

The Chr 12 locus regulates expression of 111 genes and is delineated by two SNP markers, rs13481620 at 98.47783 Mb and rs8273308 at 99.83812 Mb (Figure 4, average LRS = 42.6, genome-wide p value < 0.03 and 0.25 False Discovery Rate [28]). Surprisingly, the mean LRS for the transcripts that are cis-regulated (LRS = 31.4) at this locus is lower than that for the trans-regulated (LRS = 43.1) ones. It was hypothesized that this locus contains a gene that serves as a liver-specific "master regulator" of this Chr 12 trans-band. To identify the candidate gene, genes in a 4 Mb region (98 - 102 Mb) around the two eQTL markers identified above were considered (Figure 4, lower panel).

It was reasoned that the candidate gene that may be responsible for the variation in expression between the transcripts associated with this locus should satisfy the following properties: 1) be cis-regulated at the Chr 12 locus or contain non-synonymous coding SNPs between parental alleles, 2) have median to high expression in liver and 3) exhibit strong correlation in gene expression between the candidate gene and the trans-regulated transcripts when separated by parental allele at this locus. Five genes that are located in this region are cis-regulated: *Dicer1, Serpina3k, Serpina3n, Serpina9* and *Serpina12* (Figure 4, middle panel, genes identified in italics). Furthermore, a number of genes in this region, including *Serpina3k, Serpina3n, Serpina9,* and *Serpina12*, have known non-synonymous coding SNPs between the C57BL/6J and DBA/2J strains. *Serpina3k, Serpina3n, Dicer1, Serpina9* and *Serpina12*, among several other genes, also have median to high relative mRNA expression in liver. Lastly, when the strength of the correlation between expression of each transcript regulated by this locus and expression of all other transcripts located in the Chr 12 locus is plotted (Figure 4, top panel), it is evident that *Dicer1, Serpina3k, Serpina3n, Serpina9, Serpina12* and four other transcripts have a clear separation (positive or negative correlation) according to the parental strain allele at this locus. Thus, it appears that any of the 5 cis-regulated genes at this locus: *Dicer1, Serpina3k, Serpina3n, Serpina9, Serpina12* is likely to be the candidate "master regulator" gene in liver.

Dicer1 is a logical candidate for the Chr 12 locus since it is involved in post-transcriptional regulation of genes via cleavage of double-stranded RNAs (dsRNA) into ~20-25 base pair fragments. These fragments are then processed by the RNA-induced silencing complex, which recognizes sequences complementary to the dsRNA and prevents protein synthesis through either mRNA degradation or inhibition of translation [29]. To date, no specific gene regulatory function has been proposed for Dicer1 in the liver. To test the hypothesis that Dicer1 is a master-regulator of gene expression in the liver, expression was compared between the Chr12 locus-regulated genes in livers of wildtype and *Dicer1* heterozygous [*Dicer1* null mutation is embryonic lethal [24]] mice by quantitative real time PCR. As a negative control, a number of randomly selected genes that are not regulated by the Chr12 locus were selected. Contrary to our hypothesis, no consistent correlation was found between expression of *Dicer1* and Chr12 locus-regulated genes (data not shown) which suggests that *Dicer1* does not appear to be the master regulator at this locus. It should be noted, however, that *Dicer1* heterozygous mice may not be the most appropriate system for testing this hypothesis since *Dicer1* mRNA levels in heterozygotes are 134% of wildtype levels and a small sample size (n = 3 per group) limits the power of the analysis (p = 0.17).

Next, other means of biological interpretation of the data were considered. Gene Ontology (GO) and transcription factor binding site analyses of the Chr12 locus trans-regulated genes were performed. GOMiner [30] examination of the 111 transcripts with maximum QTLs at the Chr12 locus identified significant enrichment for a single biological process category – cell surface receptor linked signal transduction (p=8.74×10^{-4}). The genes from this category that are trans-regulated by the Chr12 locus are mainly olfactory receptors: *Bsf3, Rqcd1, Gpr50, Tcp10c, P2ry10, Olfr1403, Olfr1443, Olfr401, Olfr512, Olfr935, Olfr1341, Olfr341, Olfr656, Olfr1365, Mesp2, Met, Ltbp3, Fstl3, Centd2,* and *Rassf3.*

For the transcription factor binding site analysis, the trans-regulated transcripts at the Chr12 locus (LRS values greater than or equal to 30) were divided into two groups: those with high expression when either C57BL/6J or DBA/2J allele is present at the Chr12 locus. The National Cancer Institute's Advanced Biomedical Computing Center promoter analysis tool (grid.abcc.ncifcrf.gov/promoters.php) found no common transcription factor binding sites for the C57BL/6J list. LVc-Mo-MuLV and SV40.11 binding sites were identified as significant (p = 9.766e-04) in the DBA/2J list. The oPossum [22] tool identified Freac-2 site as significant in both C57BL/6J and DBA/2J lists (p = 6.026e-02 and p = 1.823e-02,

respectively), while ARNT (p = 8.577e-02) and SOX-9 (p = 9.159e-02) sites were also found to be common for DBA/2J allele-containing genes. The PAINT transcription factor tool [23] was also applied to the data and no significant transcription factor binding sites between the two lists were found after FDR correction of the p-values. Similarly to our observation of the lack of a consistent signal for a transcription factor, Yvert et al. [31] and Kulp et al. [32] found that the genes in trans-regulated bands are not enriched for transcription factors or biological function. This suggests that the trans-regulated genes at the Chr12 locus have a complex mechanism of regulation that is yet to be discovered.

The other candidates for the master regulator at the Chr12 locus are Serpina3-family genes. These genes are the murine orthologs of human α1-anti-chymotrypsin, a serine protease inhibitor. Serpina3n is an acute phase protein that increases 4 to 5-fold in inflammation and infection [33]. In humans, α1-anti-chymotrypsin is an inhibitor of neutrophil elastase, cathepsin G, mast cell chymase & pancreatic chymotrypsin [34]. In the mouse, *Seprin3a* targets are leukocyte elastate, cathepsin G and chymotrypsin [35]. While humans have one copy of α1-anti-chymotrypsin at 14q32, the mouse has 14 copies at 12E1 [36]. α1-Anti-chymotrypsin has been shown to be present in the amyloid plaques of Alzheimer's patients [36]. Elzouki et al. [37] found an association between low plasma α1-anti-chymotrypsin levels and propensity to contract the hepatitis B & C virus. A related protease inhibitor, serpina1 (α1-antitrypsin) is involved in emphysema due to a failure to inhibit neutrophil elastase and cirrhosis due to an accumulation of serpina1 polymers in the hepatocytes [34,38]. Although the overall structure is well conserved in the 14 member mouse *Serpina3* family, the reactive center loop is widely divergent, suggesting that these enzymes have function other than protease inhibition. Interestingly, human α1-anti-chymotrypsin was reported to be able to bind to DNA and has been found to inhibit DNA polymerase and DNA primase *in vitro*[34]. Horvath et al. [35] performed a detailed structural analysis of mouse SERPINA3N and found that it contains a DNA binding domain similar to one described in human α1-anti-chymotrypsin [39]. However, it was also reported that human α1-anti-chymotrypsin binds to double stranded DNA without specificity to known DNA binding motifs [39]. It has also been shown that some serpins may require cofactors for activation [34] which raises the possibility that the behavior of the *Serpina3* family changes depending on the environment. Collectively, while there is no firm evidence for the role of *Serpina3* genes in regulation of gene expression, we posit that our data points to the potential novel role of this family of genes in regulating liver gene expression.

eQTL analysis facilitates the discovery of novel genotype-phenotype correlations

WebQTL contains comprehensive manually curated publicly available data for phenotypic and gene expression profiling of a number of recombinant inbred and F2 crosses in both mouse and rat along with the dense genetic marker maps for these strains. Thus, this data can be used to search for correlations between phenotypes, gene expression and genetic markers, i.e., to perform *in silico* genotype-phenotype association analysis. The inherent significance of the defined reference genetic populations, such as BXD RI strains, is in the ability to connect historical data generated in many laboratories to the exact genetic map of each strain. This provides an exceptional opportunity to add value and depth to the biological interpretation of the data from model organisms. Thus, even though the BXD RI panel of strains has not been used extensively to profile liver disease-specific responses, as compared to a wealth of behavioral phenotypes published over the years, it is not unreasonable to anticipate that more data will become available soon.

Here, to illustrate the power of combining genome-wide, liver expression profiling in a reference mouse panel with phenotype profiling, we identify several phenotypes that strongly correlate with the expression of liver transcription factor hepatocyte nuclear factor 4-gamma (*Hnf4g*, located on Chr3 at 3.620141 Mb) using standard tools available in

WebQTL. *Hnf4g* is trans-regulated by the Chr12 locus (Figure 5, left panel). Several of the BXD phenotypes are also regulated by this locus and there is strong correlation (Figure 5, right panel) between *Hnf4g* expression and the induction of serum IL-6 after TNF injection, lethality due to TNF injection and decreased body temperature after TNF injection [40]. Furthermore, both the expression of *Hnf4g* and values of these phenotypes separate by parental allele at the location of the *Hnf4g* gene on proximal Chr3.

Knowledge of the variability in basal gene expression in liver as a tool for selecting relevant strains to probe biology

Genetic engineering has been a powerful and useful tool in biomedical research. However, there are a number of instances when generating a knockout or knockin mouse is neither the best option, nor it is feasible. The process is both costly and lengthy. Furthermore, some genes are embryonic lethal while others are sufficiently redundant, thus limiting the ability to generate biologically meaningful data using genetic engineering approaches. Thus, we propose that understanding of the degree of variability in gene expression between strains in a reference population of mice may be used to model the potential biological effects of naturally occurring differences in mRNA levels between individuals. In fact, it is extremely rare that people are complete nulls for a particular gene, but the polymorphisms in certain genes are known to predispose the individuals to some exposures or lead to disease [41,42]. Accordingly, the liver expression data in WebQTL may be utilized to select strains that differ in basal mRNA level of genes of interest and then used for phenotypic studies that are designed to test the role of the genes in a particular phenotype.

The flavin-containing mono-oxygenases are a class of phase I enzymes that oxidize organic nitrogen and sulfur containing compounds such as cimitidine, methimazole and nicotine [43-45]. Fmo3 is the most common isoform in the human liver [46]. While no sex differences in human activity of Fmo3 have been shown, in mice *Fmo3* expression has been shown to be much higher in females [46,47]. Basal expression of *Fmo3* varies widely across the BXD strains (Figure 6, top panel). We confirmed this by running quantitative real time PCR on strains with high, medium and low expression of *Fmo3* in 9 out of 40 strains (Figure 6, lower panel). The correlation between the expression measured by microarray and PCR is strong ($R^2 = 0.74$). Thus, WebQTL can be used to query the expression database to find genes with the highest differential expression across strains to generate a potential list of candidate strains for hypothesis testing. Since *Fmo3* expression varies widely across BXD strains and mutations in this gene have been implicated in trimethylaminuria [41], a disease in which trimethylamine is not metabolized but is excreted in the breath and sweat, leading to a persistent fishy body odor, we suggest that BXD strains with high or low relative expression of *Fmo3* may be used to model this disease. Similar logic may be used to test other genes without the cost and time of generating knockout animals.

In conclusion, this study describes a new public resource that will facilitate our understanding of the genetic regulation of gene expression in liver. We describe several genetic loci that control expression of large numbers of genes. By using eQTL mapping, we identified the *Serpina3* family of genes as potential novel master regulators of transcription in the liver. By comparing the liver and brain transcriptome maps, we highlighted tissue-specific differences in regulation of gene expression. Finally, this study demonstrates how this publicly available data may be used to infer genotype-phenotype correlations, generate testable hypothesis, and select mouse strains for further testing based on the genetically-determined differences of expression of the key genes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. Science. 2002; 296:752–755. [PubMed: 11923494]

2. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, et al. Genetics of gene expression surveyed in maize, mouse and man. Nature. 2003; 422:297–302. [PubMed: 12646919]

3. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, et al. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat Genet. 2005; 37:233–242. [PubMed: 15711545]

4. Waring JF, Jolly RA, Ciurlionis R, Lum PY, Praestgaard JT, Morfitt DC, et al. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. Toxicol Appl Pharmacol. 2001; 175:28–42. [PubMed: 11509024]

5. Powell CL, Kosyk O, Ross PK, Schoonhoven R, Boysen G, Swenberg JA, et al. Phenotypic anchoring of acetaminophen-induced oxidative stress with gene expression profiles in rat liver. Toxicol Sci. 2006; 93:213–222. [PubMed: 16751229]

6. Taylor, BA. Recombinant inbred strains. In: Lyon, ML.; Searle, AG., editors. Genetic Variants and Strains of the Laboratory Mouse. Oxford, UK: Oxford University Press; 1989. p. 773-796.

7. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989; 121:185–199. [PubMed: 2563713]

8. Farrall M. Quantitative genetic variation: a post-modern view. Hum Mol Genet. 2004; 13(Spec No 1):R1–R7. [PubMed: 14962979]

9. Morse, HCI., editor. Origins of inbred mice: Proceedings of a workshop. Vol. 78. Academic Press; Feb 14. 1978

10. Peirce JL, Lu L, Gu J, Silver LM, Williams RW. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC Genet. 2004; 5:7. [PubMed: 15117419]

11. Treadwell JA. Integrative strategies to identify candidate genes in rodent models of human alcoholism. Genome. 2006; 49:1–7. [PubMed: 16462896]

12. Gill K, Liu Y, Deitrich RA. Voluntary alcohol consumption in BXD recombinant inbred mice: relationship to alcohol metabolism. Alcohol Clin Exp Res. 1996; 20:185–190. [PubMed: 8651451]

13. Bigelow SW, Nebert DW. The murine aromatic hydrocarbon responsiveness locus: a comparison of receptor levels and several inducible enzyme activities among recombinant inbred lines. J Biochem Toxicol. 1986; 1:1–14. [PubMed: 2856069]

14. Lee GH, Bennett LM, Carabeo RA, Drinkwater NR. Identification of hepatocarcinogen-resistance genes in DBA/2 mice. Genetics. 1995; 139:387–395. [PubMed: 7705639]

15. Davis RC, Schadt EE, Cervino AC, Peterfy M, Lusis AJ. Ultrafine mapping of SNPs from mouse strains C57BL/6J, DBA/2J, and C57BLKS/J for loci contributing to diabetes and atherosclerosis susceptibility. Diabetes. 2005; 54:1191–1199. [PubMed: 15793261]

16. Nishina PM, Wang J, Toyofuku W, Kuypers FA, Ishida BY, Paigen B. Atherosclerosis and plasma and liver lipids in nine inbred strains of mice. Lipids. 1993; 28:599–605. [PubMed: 8355588]

17. Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T, et al. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. Nat Genet. 2005; 37:225–232. [PubMed: 15711547]

18. Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, et al. Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods. 2005; 2:351–356. [PubMed: 15846362]

19. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 2002; 30:e15. [PubMed: 11842121]

20. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994; 138:963–971. [PubMed: 7851788]

21. Storey JD. A direct approach to false discovery rates. J R Statist Soc B. 2002; 64:479–498.

22. Ho Sui SJ, Mortimer JR, Arenillas DJ, Brumm J, Walsh CJ, Kennedy BP, et al. oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. Nucleic Acids Res. 2005; 33:3154–3164. [PubMed: 15933209]

23. Vadigepalli R, Chakravarthula P, Zak DE, Schwaber JS, Gonye GE. PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. OMICS. 2003; 7:235–252. [PubMed: 14583114]

24. Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, et al. Dicer is essential for mouse development. Nat Genet. 2003; 35:215–217. [PubMed: 14528307]

25. Doss S, Schadt EE, Drake TA, Lusis AJ. Cis-acting expression quantitative trait loci in mice. Genome Res. 2005; 15:681–691. [PubMed: 15837804]

26. Peirce JL, Li H, Wang J, Manly KF, Hitzemann RJ, Belknap JK, et al. How replicable are mRNA expression QTL? Mamm Genome. 2006; 17:643–656. [PubMed: 16783644]

27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]

28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B. 1995; 57:289–300.

29. Hannon GJ. RNA interference. Nature. 2002; 418:244–251. [PubMed: 12110901]

30. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol. 2003; 4:R28. [PubMed: 12702209]

31. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, et al. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet. 2003; 35:57–64. [PubMed: 12897782]

32. Kulp DC, Jagalur M. Causal inference of regulator-target pairs by gene mapping of expression phenotypes. BMC Genomics. 2006; 7:125. [PubMed: 16719927]

33. Kalsheker N, Morley S, Morgan K. Gene regulation of the serine proteinase inhibitors alpha1-antitrypsin and alpha1-antichymotrypsin. Biochem Soc Trans. 2002; 30:93–98. [PubMed: 12023832]

34. Janciauskiene S. Conformational properties of serine proteinase inhibitors (serpins) confer multiple pathophysiological roles. Biochim Biophys Acta. 2001; 1535:221–235. [PubMed: 11278163]

35. Horvath AJ, Irving JA, Rossjohn J, Law RH, Bottomley SP, Quinsey NS, et al. The murine orthologue of human antichymotrypsin: a structural paradigm for clade A3 serpins. J Biol Chem. 2005; 280:43168–43178. [PubMed: 16141197]

36. Forsyth S, Horvath A, Coughlin P. A review and comparison of the murine alpha1-antitrypsin and alpha1-antichymotrypsin multigene clusters with the human clade A serpins. Genomics. 2003; 81:336–345. [PubMed: 12659817]

37. Elzouki AN, Verbaan H, Lindgren S, Widell A, Carlson J, Eriksson S. Serine protease inhibitors in patients with chronic viral hepatitis. J Hepatol. 1997; 27:42–48. [PubMed: 9252072]

38. Perlmutter DH. Pathogenesis of chronic liver injury and hepatocellular carcinoma in alpha-1-antitrypsin deficiency. Pediatr Res. 2006; 60:233–238. [PubMed: 16864711]

39. Naidoo N, Cooperman BS, Wang ZM, Liu XZ, Rubin H. Identification of lysines within alpha 1-antichymotrypsin important for DNA binding. An unusual combination of DNA-binding elements. J Biol Chem. 1995; 270:14548–14555. [PubMed: 7782318]

40. Libert C, Wielockx B, Hammond GL, Brouckaert P, Fiers W, Elliott RW. Identification of a locus on distal mouse chromosome 12 that controls resistance to tumor necrosis factor-induced lethal shock. Genomics. 1999; 55:284–289. [PubMed: 10049582]

41. Dolphin CT, Janmohamed A, Smith RL, Shephard EA, Phillips IR. Missense mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour syndrome. Nat Genet. 1997; 17:491–494. [PubMed: 9398858]

42. Rudnick DA, Perlmutter DH. Alpha-1-antitrypsin deficiency: a new paradigm for hepatocellular carcinoma in genetic liver disease. Hepatology. 2005; 42:514–521. [PubMed: 16044402]

43. Ziegler DM. Recent studies on the structure and function of multisubstrate flavin-containing monooxygenases. Annu Rev Pharmacol Toxicol. 1993; 33:179–199. [PubMed: 8494339]

44. Cashman JR, Park SB, Berkman CE, Cashman LE. Role of hepatic flavin-containing monooxygenase 3 in drug and chemical metabolism in adult humans. Chem Biol Interact. 1995; 96:33–46. [PubMed: 7720103]

45. Ripp SL, Overby LH, Philpot RM, Elfarra AA. Oxidation of cysteine S-conjugates by rabbit liver microsomes and cDNA-expressed flavin-containing mono-oxygenases: studies with S-(1,2-dichlorovinyl)-L-cysteine, S-(1,2,2-trichlorovinyl)-L-cysteine, S-allyl-L-cysteine, and S-benzyl-L-cysteine. Mol Pharmacol. 1997; 51:507–515. [PubMed: 9058607]

46. Ripp SL, Itagaki K, Philpot RM, Elfarra AA. Species and sex differences in expression of flavin-containing monooxygenase form 3 in liver and kidney microsomes. Drug Metab Dispos. 1999; 27:46–52. [PubMed: 9884308]

47. Falls JG, Blake BL, Cao Y, Levi PE, Hodgson E. Gender differences in hepatic expression of flavin-containing monooxygenase isoforms (FMO1, FMO3, and FMO5) in mice. J Biochem Toxicol. 1995; 10:171–177. [PubMed: 7473608]

Abbreviations

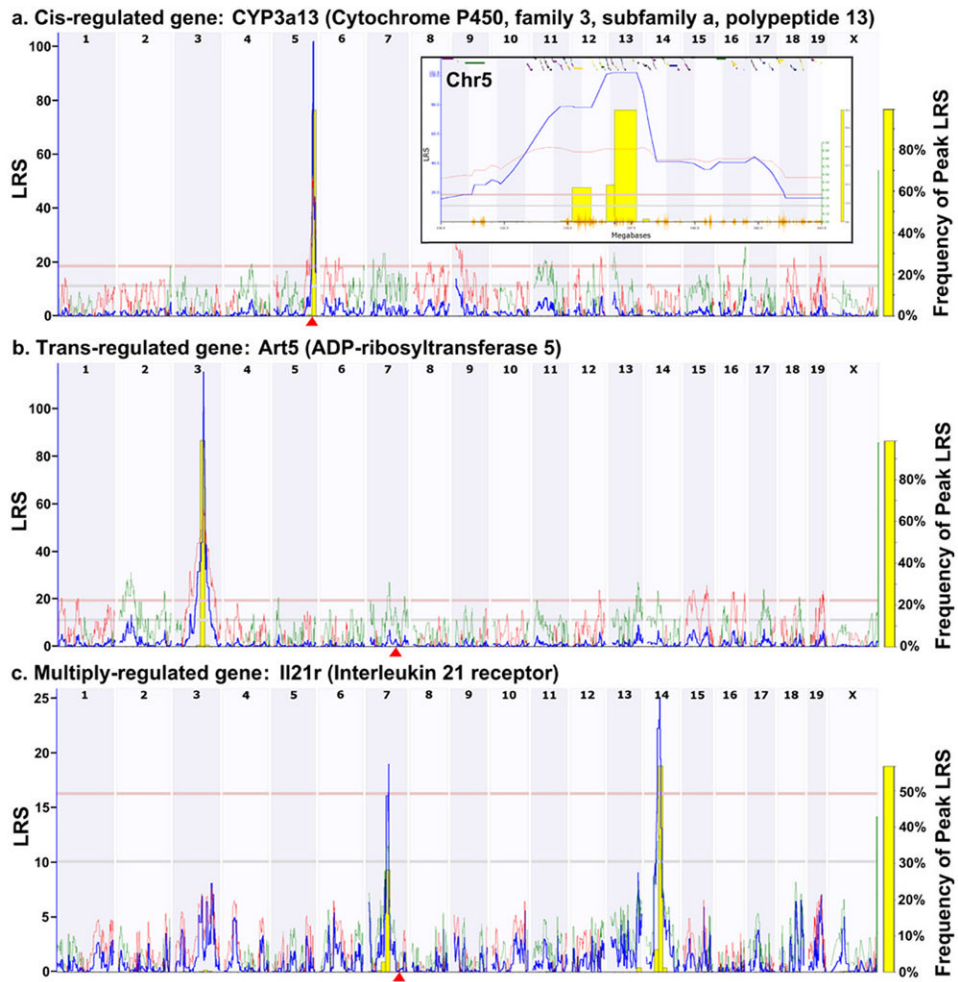|	
AIL	Advanced intercross line
FDR	False discovery rate
GO	Gene Ontology
QTL	Quantitative trait loci
RI	Recombinant inbred
LRS	Likelihood ratio statistic

Figure 1. WebQTL interval mapping reveals genetic control of gene expression
a, An example of a cis-regulated gene (*Cyp3a13*) where the QTL is co-located with the gene. The horizontal axis displays the mouse genome. The vertical axis displays the Likelihood Ratio Statistic (LRS), The gene location is shown by the red triangle. The red horizontal line indicates a significant level of association as determined by permutation analysis. The grey horizontal line below it represents a suggestive association. The blue line displays the LRS along the genome and the yellow bars are the results of a bootstrap analysis. The inset shows a zoomed-in view of the *Cyp3a13* QTL on Chr5 (130-145Mb) and illustrates the features of WebQTL. The red line indicates that C57BL/6J alleles at the marker increase expression of the gene. The colored boxes along the top represent known genes which, when selected, will take the user to the NCBI EntrezGene entry for that gene. A SNP track is displayed along the bottom as an orange seismogram. **b**, An example a trans-regulated gene (*Art5*). **c**, An example of a gene (*Il21r*) that is regulated by multiple loci.
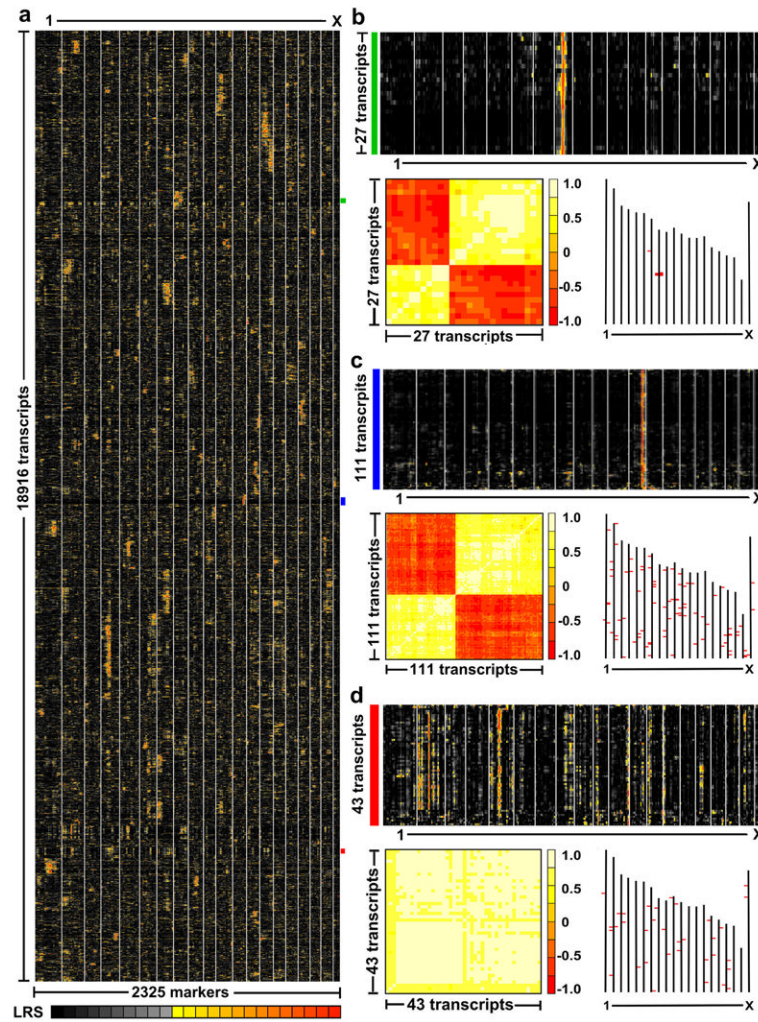
Figure 2. Genome-wide clustering of the genetic control of gene expression in liver
a, Hierarchical clustering diagram of all transcripts on the array by Likelihood Ratio Statistic (LRS) profile. Rows – transcripts on the microarray, columns – SNP markers used for the QTL analysis. Strength of the LRS values is depicted using a head map with black being the lowest LRS, and bright red – the highest. **b**, A zoomed-in view of the cluster of genes controlled by a single genetic locus on Chr8. An auto-correlation matrix of the measured expression values for the transcripts in the cluster and a plot of chromosomal location of these transcripts are shown below. **c**, A cluster of 111 transcripts controlled by one locus on Chr12. **d**, A cluster of 43 transcripts that are controlled by a complex set of loci on several chromosomes.
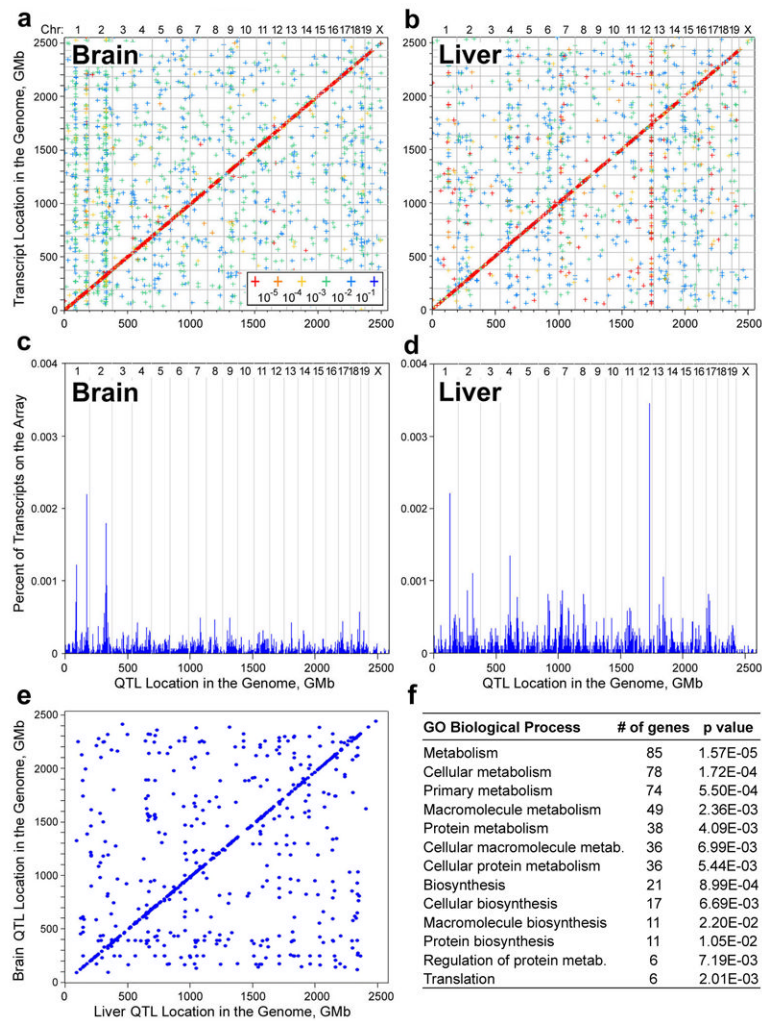
Figure 3. Tissue-specific transcriptome maps reveal differences and similarities in genetic regulation of gene expression
a-b, The brain [26] and liver (this study) transcriptome maps. The horizontal axis shows the genomic location of each genetic marker along the genome. The vertical axis shows the genomic location of each transcript probed by the microarray analysis. Each cross represents the location of the maximum QTL for a particular gene. Cis-regulated genes, where the QTL is co-located with the gene, fall along a 45 degree line. The vertical lines correspond to a locus with strong trans-control over many genes. The major loci of control differ markedly in the murine brain and liver. **c-d**, Histograms counting the proportion of transcripts on the array regulated at each marker. **e**, A comparative transcriptome map between the murine liver and brain. Significant QTLs in the liver and brain are plotted along the horizontal and vertical axes, respectively. Points along the diagonal represent transcripts whose maximum QTL is the same in both tissues, indicating a similar mechanism of expression control. Points plotted off of the diagonal represent transcripts that are controlled by different loci in the two tissues. **f**, Gene Ontology analysis of the significantly enriched biological processes common between liver and brain in the mouse
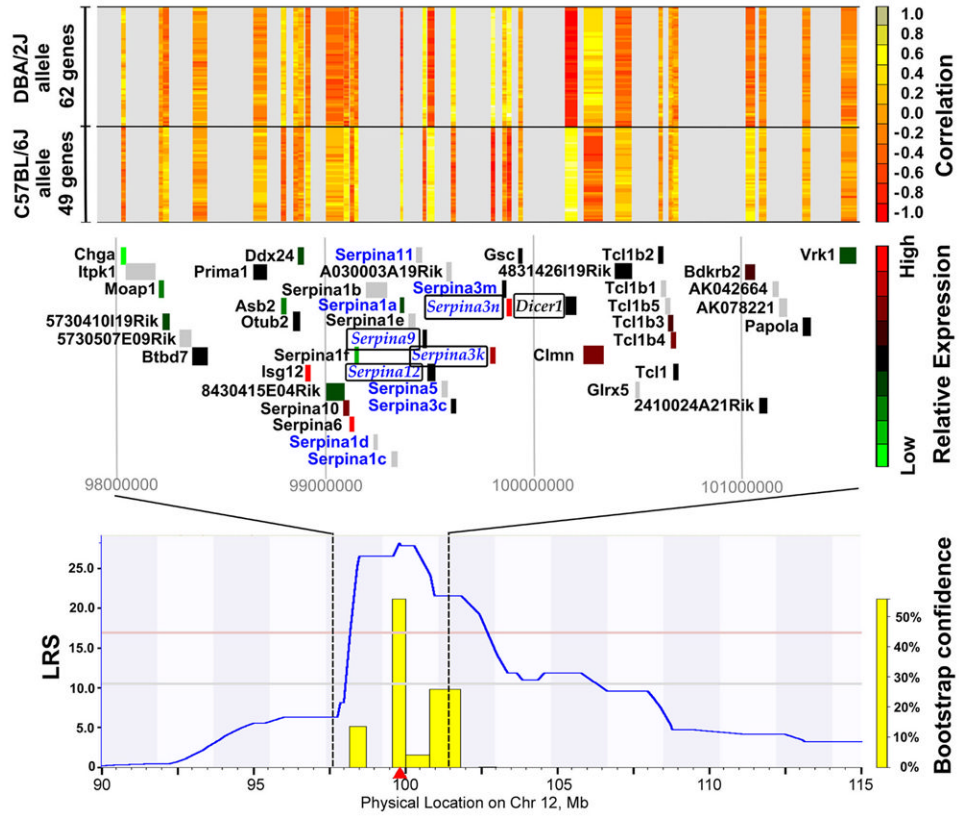
Figure 4. Chr12 locus is a master-regulator of gene expression in mouse liver
A zoomed-in view of the region on Chr12 that controls a large number of trans-regulated transcripts is shown in the bottom panel. The middle panel is a detail of the QTL peak from 98 Mb to 102 Mb showing the genes in this region. Genes labeled in blue text have non-synonymous coding SNPs. Genes in italics are cis-regulated. Genes in plain text are trans-regulated. The gene location marker is color coded based on relative expression in the mouse liver. Green indicates low expression; red indicates high expression; grey indicates that the gene is not represented on the microarray. The top panel shows the correlation between the expression of each gene located in this locus and the putative trans-regulated genes. Red indicates negative correlation; yellow indicates positive correlation; grey indicates no information. Genes in the upper part of the panel are highly expressed in strains with the DBA/2J allele at this Chr12 locus. Genes on the bottom of the panel are highly expressed in strains with the C57BL/6J allele is at this locus.
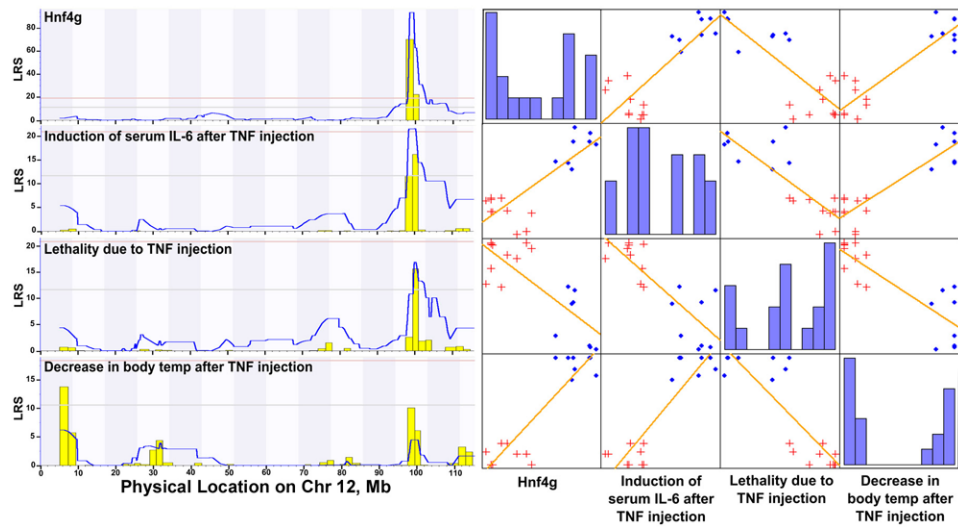
Figure 5. *In silico* **discovery of gene expression to phenotype correlations using WebQTL**
The left panel shows QTL interval maps on Chr12 for hepatocyte nuclear factor 4, gamma (*Hnf4g*) gene and three related phenotypes. The right panel shows correlation pair-wise correlation plots for *Hnf4g* expression and the phenotypes. Each dot represents the measurements for one strain. Blue dots indicate strains in which the C57BL/6J allele is present at the Chr12 (99.83Mb) locus. Red crosses represent strains with the DBA/2J allele at this locus. The blue histograms along the diagonal represent the distribution of all values of the phenotype.
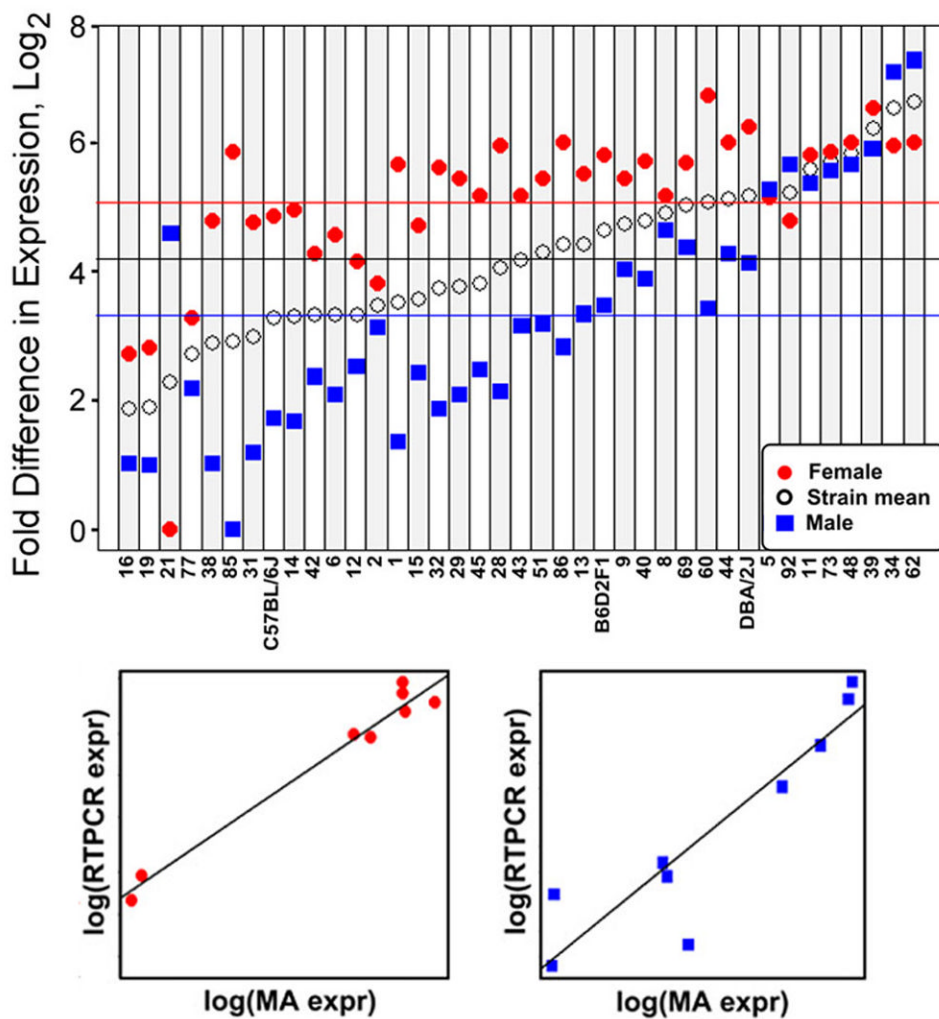
Figure 6. WebQTL-assisted strain selection for phenotypic profiling
The main panel shows relative expression of *Fmo3* (flavin monooxygenase 3) across BXD
strains measured by microarray. Red circles are values in females; blue squares – in males
and the strain means are shown as black circles. The lower panels show relative *Fmo3*
expression in select strains as measured by microarray vs. quantitative real time PCR. The
expression of *Fmo3* in BXD21 strain was set to 0.