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,[3,4] Kenneth F. Manly,[3] Robert W. Williams,[3] Andy Perkins,[5] Michael A. Langston,[5] David W. Threadgill,[6]† and Ivan Rusyn[1]†

The liver is the primary site for the metabolism of nutrients, drugs, and chemical agents. Although metabolic pathways are complex and tightly regulated, genetic variation among individuals, reflected in variations in gene expression levels, introduces complexity into research on liver disease. This study dissected genetic networks that control liver gene expression through the combination of large-scale quantitative mRNA expression analysis with genetic mapping in a reference population of BXD recombinant inbred mouse strains for which extensive single-nucleotide polymorphism, haplotype, and phenotypic data are 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. These data were used to map quantitative trait loci (QTLs) responsible for variations in the expression of about 19,000 transcripts. We identified polymorphic local and distant QTLs, including several loci that control the expression of large numbers of genes in liver, by comparing the physical transcript position with the location of the controlling QTL. *Conclusion:* The data are available through a public web-based resource (www.genenetwork.org) that allows custom data mining, identification of coregulated transcripts and correlated phenotypes, cross-tissue, and cross-species comparisons, as well as testing of a broad array of hypotheses. (HEPATOLOGY 2007;46:548-557.)

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 the 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 the transcript expression to phenotypes. However, they do not lead to detailed gene expression networks in which the expression of 1 gene is found to control the expression of another.

Recombinant inbred (RI) mice are created by the crossing of 2 parental strains followed by sib-mating for over 20 generations.[6] Strains created in this way have the advantage of being homozygous at almost every location along the genome. 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 with quantitative trait locus (QTL) mapping techniques.[7] The relationships between the phenotypes and genotypes are calculated with 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 called expression quantitative trait locus (eQTL) mapping.[2,8]

The BXD panel of RI strains was created from the C57BL/6J and DBA/2J parental strains.[9] At each locus,

Abbreviations: eQTL, expression quantitative trait locus; FDR, false discovery rate; Fmo3, flavin monooxygenase 3; GO, gene ontology; Hnf4g, hepatocyte nuclear factor 4-gamma; Il21r, interleukin 21 receptor; QTL, quantitative trait locus; LRS, likelihood ratio statistic; RI, recombinant inbred; SDP, strain distribution pattern; SNP, single-nucleotide polymorphism; TNF, tumor necrosis factor.

From the [1]Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; the [2]Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; the [3]Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, TN; the [4]Key Laboratory of Nerve Regeneration, Nantong University, Nantong, People's Republic of China; [5]Department of Computer Science, University of Tennessee, Knoxville, TN; and the [6]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Received December 8, 2006; accepted January 31, 2007.

Supported in part by grants and contracts from the National Institutes of Health (ES011391, ES010126, AA016258, AA013499, AA013513, and AA014425), the Environmental Protection Agency (STAR-RD832720), and the Bioinformatics and Computational Biology Training Program at the University of North Carolina at Chapel Hill (D.G.).

**These authors contributed equally to this work.*

†Equally contributing senior author.

Address reprint requests to: 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. E-mail: iir@unc.edu; fax: 919-843-2596.

Published online in Wiley InterScience (www.interscience.wiley.com).

DOI 10.1002/hep.21682

Potential conflict of interest: Nothing to report.

Supplementary material for this article can be found on the HEPATOLOGY website (http://interscience.wiley.com/jpages/0270-9139/suppmat/index.html).

each strain will contain an allele from either the C57BL/6J parent or the DBA/2J parent. These 2 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 the elucidation of the genetic control of certain diseases. For example, the BXD lines have been used to study alcohol preference and tolerance,[10] alcohol metabolism,[11] responsiveness to aromatic hydrocarbons,[12] *N,N*-diethylnitrosamine–induced hepatocarcinogenesis,[13] diabetes and atherosclerosis.[14,15] Recently, BXD mice were used for eQTL studies that elucidated the genetics of gene expression in the brain[3] and hematopoietic stem cells.[16]

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 our 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, we identified potential key regulators of gene expression in the liver. The comparison of the liver and brain transcriptome maps demonstrated tissue-specific differences in the regulation of gene expression. Finally, we demonstrate how the data collected in this study may be used to infer genotype-phenotype correlations, generate a testable hypothesis, and identify strains that may differ in responsiveness to xenobiotics based on the genetically determined differences in the expression of the key genes.

Materials and Methods

Animals and Tissues. The BXD1 through BXD42 mice are original RI strains available from the Jackson Laboratory (Bar Harbor, ME). The BXD43 through BXD100 lines were generated with ninth-generation or tenth-generation advanced intercross line progenitors.[9] Strain details are provided in Supplementary Table 1. The housing and tissue collection are detailed in the supplementary material. All animal studies for this project were approved by the Animal Care and Use Committee at the University of Tennessee at Memphis.

RNA Isolation. The total RNA was isolated from liver samples (~30 mg) with an RNeasy mini kit (Qiagen, Valencia, CA) as detailed by the manufacturer. For each microarray, RNA was pooled from 2-3 mice of the same sex and strain. One array per sex per strain was used except for strains 23, 34, and 33 (see Supplementary Table 1). The ratio of males to females was approximately 1.

Microarray Analysis. A balanced design was used to assign samples into 8 batches. Each batch of arrays (Agilent mouse oligo microarray, G4121) was balanced by sex and strain and contained both interbatch and intrabatch technical replicates. Interbatch normalization was carried out with a nested analysis of variance mixed model, with samples within each batch crossed with sex and strain. Additional details are provided in the supplementary material. The microarray data are available at http://genome.unc.edu.

QTL Analysis and WebQTL (www.genenetwork.org). In the BXD strains, 3795 single-nucleotide polymorphisms (SNPs; mean spacing of 0.6 centimorgan or 0.7 megabase [Mb]) flank regions containing unique strain distribution patterns (SDPs). A subset of 2325 SNPs were used here as haplotype markers. These were selected by the choice of 1 of the flanking markers for each unique SDP. Further information can be found in the supplementary material and at www.genenetwork.org/dbdoc/BXDGeno.html. QTL linkage mapping was carried out with the QTL Reaper software package (qtlreaper.sourceforge.net). One thousand permutations of the strain labels were performed to estimate the genome-wide *P* value.[17] Liver expression data were deposited in WebQTL. WebQTL was used to produce interval maps.

Transcriptome Map and Whole Genome QTL Clustering. The transcriptome map of the liver was produced in R with the output data from QTL Reaper, which consisted of the maximum LRS value for each transcript on the microarray and a permutation-derived *P* value (number of permutations = 1000). The *P* value threshold of 0.03 was applied at a 25% false discovery rate (FDR).[18] The distance metric used for complete (maximum distance) hierarchical clustering was 1-Pearson correlation, and it was colored by a 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),[19] and PAINT.[20] Additional information can be found in the supplementary material.

Quantitative Real Time PCR Analysis. cDNA was produced from 20 μg of RNA with the Applied Biosystems, Inc. (Foster City, CA), high-capacity cDNA archive kit. A Stratagene (La Jolla, CA) FullVelocity QYBR Green QPC Master Mix and a Stratagene Mx3000P instrument were used.

Results and Discussion

Genetic Control of Gene Expression in Liver. With the web-based eQTL mapping tools and data collected in this study, regulatory loci controlling each liver transcript

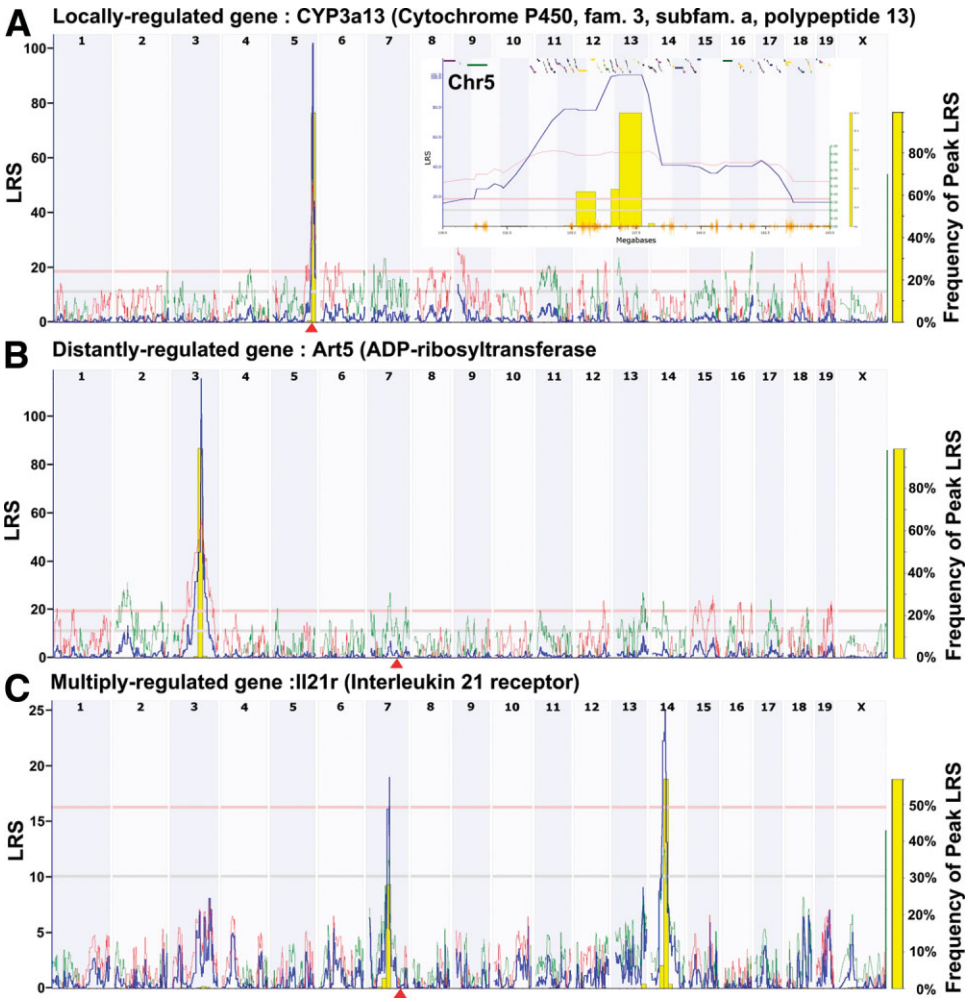

Fig. 1. WebQTL interval mapping reveals genetic control of gene expression. (A) An example of a locally regulated gene (*Cyp3a13*) in which the QTL is colocated with the gene. The horizontal axis shows the mouse genome; the vertical axis shows the LRS. The red triangle indicates the gene location. Significant and suggestive levels of association are indicated by red and gray lines, respectively. The blue line displays the LRS along the genome, and the yellow bars are the results of a bootstrap analysis. The inset is a zoomed-in view of the *Cyp3a13* QTL on Chr5 (130-145 Mb). 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. An SNP density plot is displayed along the bottom in orange. (B) An example of a distantly regulated gene (*Art5*). (C) An example of a gene (*Il21r*) that is regulated by multiple loci.

can be easily visualized in WebQTL. Figure 1 shows examples of the 3 types of expression control found in the liver. A transcript with a maximum locus of control near (±5 Mb) the genomic location of the transcript itself is considered to be a local QTL.[21,22] This implies some mechanism of control near the gene itself, perhaps a polymorphism in the promoter region. An example of a strong local QTL transcript is the cytochrome P450, family 3, subfamily a, polypeptide 13 (*Cyp3a13*) gene located on distal chromosome 5 (Fig. 1A). A transcript with a maximum locus of control far from the gene location is distantly regulated.[22] This implies regulation by another gene, perhaps a transcription factor. The ADP-ribosyltransferase 5 (*Art5*) gene, located in the middle of chromosome 7, is an example of a gene with a distant QTL whose expression levels are determined by polymorphisms on chromosome 3 (Chr3) (Fig. 1B). Many liver-expressed transcripts are regulated by multiple loci. The interleukin 21 receptor (*Il21r*) gene, located on distal chromosome 7, is a good example (Fig. 1C). *Il21r* has 2 significant QTL peaks: 1 on chromosome 7 proximal to the gene itself and 1 on chromosome 14. Although the implication is that locally regulated transcripts are regulated by cis-acting mechanisms and distantly regulated transcripts are regulated by trans-acting mechanisms, the existence of these QTLs is not sufficient to draw firm conclusions about the molecular mechanism of regulation.

To visualize the patterns of genetic control of gene expression on a genome-wide level, the 18716 annotated transcripts were clustered with the LRS vector for each transcript (Fig. 2A). The majority of the transcripts in liver are independently regulated; however, several distinct patterns emerge. There are a number of clusters of transcripts that all share a common maximal QTL and clusters that are coregulated by a complex set of common loci. We refer to clusters regulated by only 1 strong QTL as simple QTL clusters and to 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, Fig. 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

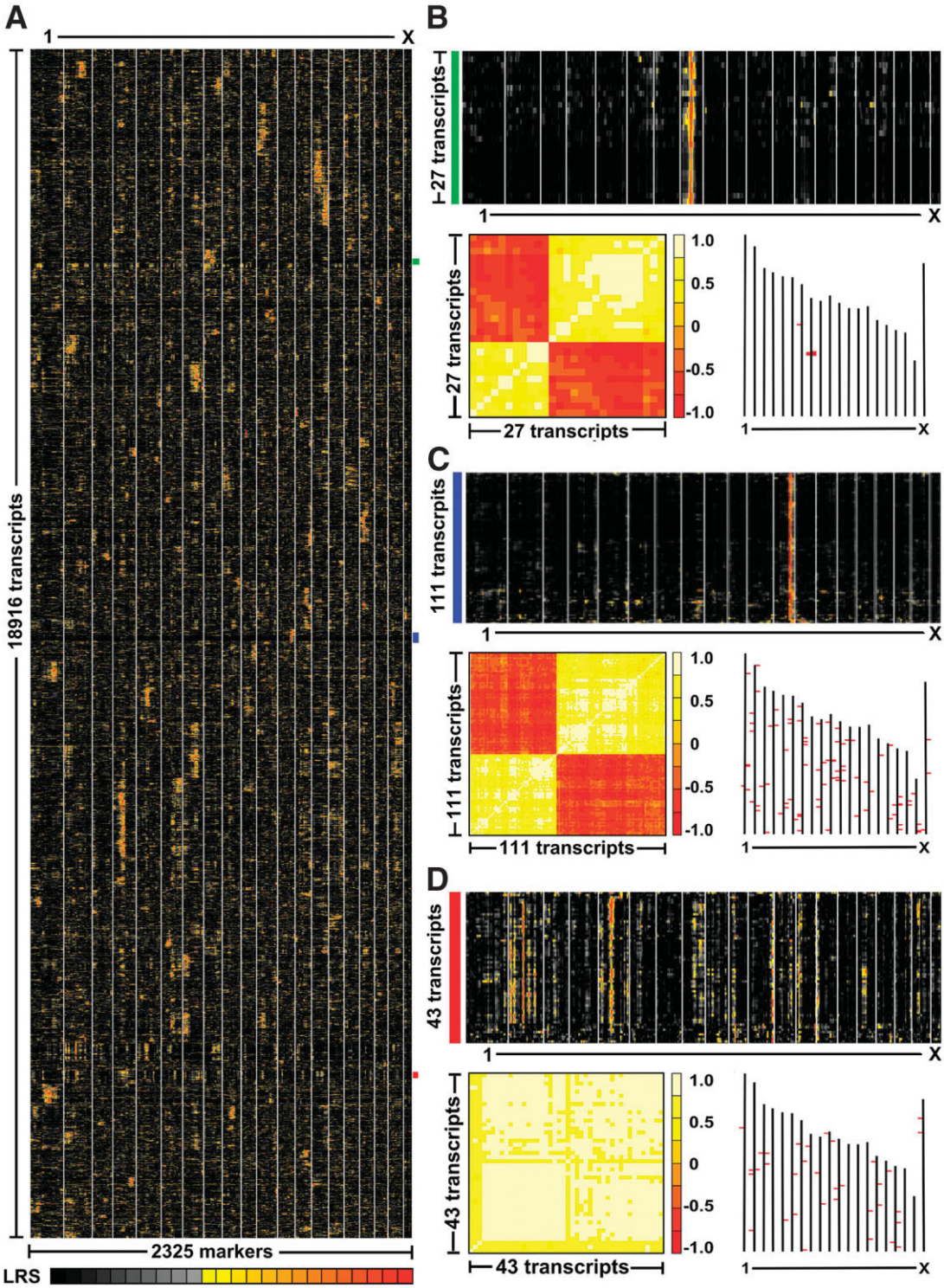

Fig. 2. Genome-wide clustering of the genetic control of gene expression in liver. (A) Hierarchical clustering diagram of all transcripts on the array by the LRS profile. The rows show the transcripts on the microarray, and the columns show the SNP markers used for the QTL analysis. The strength of the LRS values is depicted with a map, with black being the lowest LRS and bright red being the highest. (B) A zoomed-in view of the cluster of genes controlled by a single genetic locus on Chr8. An autocorrelation matrix of the measured expression values for the transcripts in the cluster and a plot of the chromosomal locations of these transcripts are shown below. (C) A cluster of 111 transcripts controlled by 1 locus on Chr12. (D) A cluster of 43 transcripts that are controlled by a complex set of loci on several chromosomes.

predominantly locally regulated genes or perhaps is due to a strain-specific difference in a regional transcriptional enhancer. Interestingly, the presence of 1 of the parental (C57BL/6J or DBA/2J) alleles at this locus strongly affects the expression of these genes (Fig. 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 (Fig. 2C). However, only 5 of the 111 transcripts in this cluster are locally regulated and hence located under the maximal QTL. This region in the genome is a master regulator because the majority of the transcripts are distantly regulated by this locus. Again, the correlation between the expression levels of these transcripts is strongly dependent on the type of 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 locally and distantly regulated transcripts in simple clusters, 31 simple clusters composed of at least 80% locally regulated transcripts were examined. It was hypothesized that the clustering of these QTLs might be due to higher gene or SNP density as opposed to biological pathways. When the gene density at each region was considered versus the adjacent upstream and downstream regions, it was found that the QTL cluster regions had 79.5 ± 70.0 (mean ± SD) genes, whereas the immediately adjacent upstream and downstream 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 upstream and downstream 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 locally regulated clusters do not appear to be enriched for any particular gene ontology (GO) category. This indicates that most of the coregulation of locally controlled genes throughout the genome is due not to functional relatedness but rather to the high gene/SNP density in each region.

Lastly, a cluster of 43 transcripts (Fig. 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 throughout the genome. The pairwise gene expression correlation matrix for these transcripts shows that mRNA levels for these genes show a high positive correlation, regardless of the allele present at each QTL.

The proportion of locally regulated eQTLs that had higher expression when the C57BL/6J allele was present was assessed. Following the criterion set out by Pierce et al.,[23] a local eQTL was defined as a transcript that has a maximum QTL within ±5 Mb of the transcript's location in the genome. At a genome-wide P value of 0.05, 1255 local eQTLs with 54.3% were found to have C57BL/6J high expression. At higher levels of statistical stringency, 1075 local eQTLs ($P = 0.01$) were found with 53.9% higher for the C57BL/6J allele, and 867 local eQTLs ($P = 0.001$) were found with 53.3% higher for the C57BL/6J allele. The data presented by Doss et al.[21] and the data from this study both show only a slight (54%-56%) enrichment for C57BL/6J high local 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 data from Pierce et al.[23] data could be attributed to low-fidelity binding of short oligonucleotides 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 the Genetic Regulation of Gene Expression. Several recent reports identified a number of master-regulatory loci in other mouse tissues.[3,16,23] Here, the mouse brain (forebrain) and liver transcriptome maps (Fig. 3) are compared to uncover the similarities and differences in the genetic regulation of gene expression across tissues. Both the brain and liver contain genes that show strong local or distant regulation at single loci or are regulated by multiple loci (Fig. 3A,B, respectively). In the brain transcriptome, 3 distinct master-regulator bands are located on chromosomes 1 and 2 (Fig. 3C). In the liver transcriptome, the strongest band is located on distal chromosome 12 (Fig. 3D), a locus that does not appear to be controlling expression of a large number of genes in the brain. Although the liver and brain both have a distant regulation band on chromosome 2 near 125 Mb, the 2 bands are not coincident. The liver band lies at 119 MB, and the brain band lies at 135 Mb; there is a difference of 16 Mb.

Next, the QTL data for the liver and brain were filtered by the selection of only those transcripts with a genome-wide P value less than or equal to 0.05.[23] This analysis identified 743 transcripts that have significant QTLs in both tissues (Fig. 3E). In both the liver and brain, 209 transcripts are regulated by the same genomic region (genes on a diagonal in Fig. 3E). A GO-based[24] analysis of the biological processes that are significantly over-represented (Fisher's exact test, $P < 0.05$ and 25% FDR)

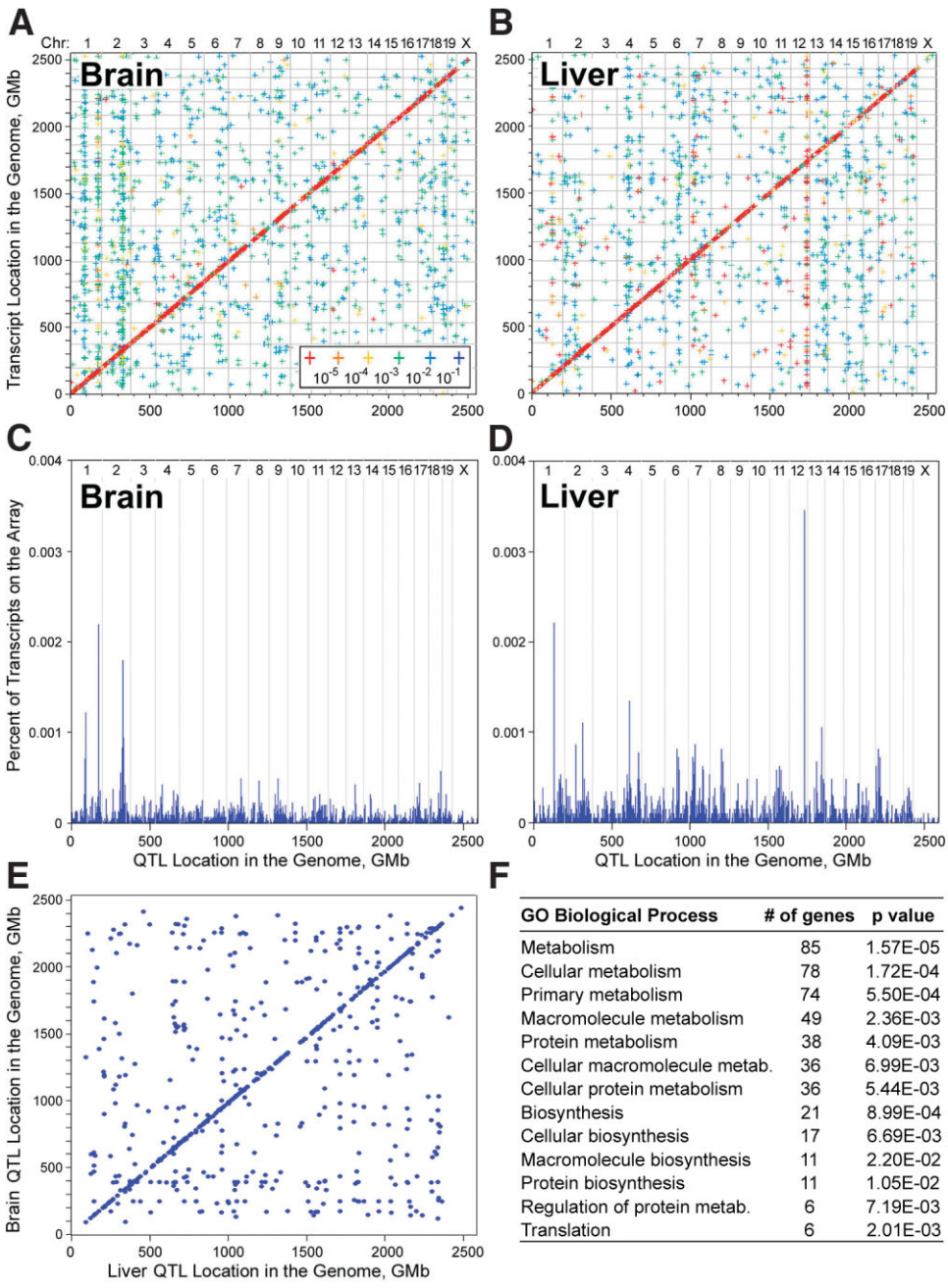

Fig. 3. Tissue-specific transcriptome maps reveal differences and similarities in the genetic regulation of gene expression. (A,B) The brain[23] and liver (this study) transcriptome maps. Genomic locations of each genetic marker (horizontal axis) or each transcript (vertical axis) along the genome are shown. Each cross represents the location of the maximum QTL for a particular gene. Locally regulated genes, in which the QTL is colocated with the gene, fall along a 45-degree line. The vertical lines correspond to a locus with strong distant control over many genes. (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. (F) GO analysis of the significantly enriched biological processes common between the liver and brain in the mouse.

among these genes identified a number of significant categories that are coregulated in both tissues (Fig. 3F). Furthermore, among the 743 transcripts, the proportion of locally 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 locations between the 2 data sets. Of the 209 transcripts that have the same QTL in the liver and the brain, about 95% are locally regulated. Collectively, this comparison indicates that important tissue-specific patterns of genetic control of gene expression can be elucidated by this approach and can potentially form the basis for comparative analyses between tissues.

Chromosome 12 Contains a Strong Liver-Specific Master-Regulatory Locus. The Chr12 locus regulates the expression of 111 genes and is delineated by 2 SNP markers, rs13481620 at 98.47783 Mb and rs8273308 at 99.83812 Mb (Fig. 4, average LRS = 42.6, genome-wide $P < 0.03$, and 0.25 FDR[25]). Surprisingly, the mean LRS for the transcripts that are locally regulated (LRS = 31.4) at this locus is lower than that for the distantly regulated ones (LRS = 43.1). It was hypothesized that this locus contains a gene that serves as a liver-specific master regulator of this Chr12 distantly regulated band. To identify the candidate gene, genes in a 4-Mb region (98-102 Mb)

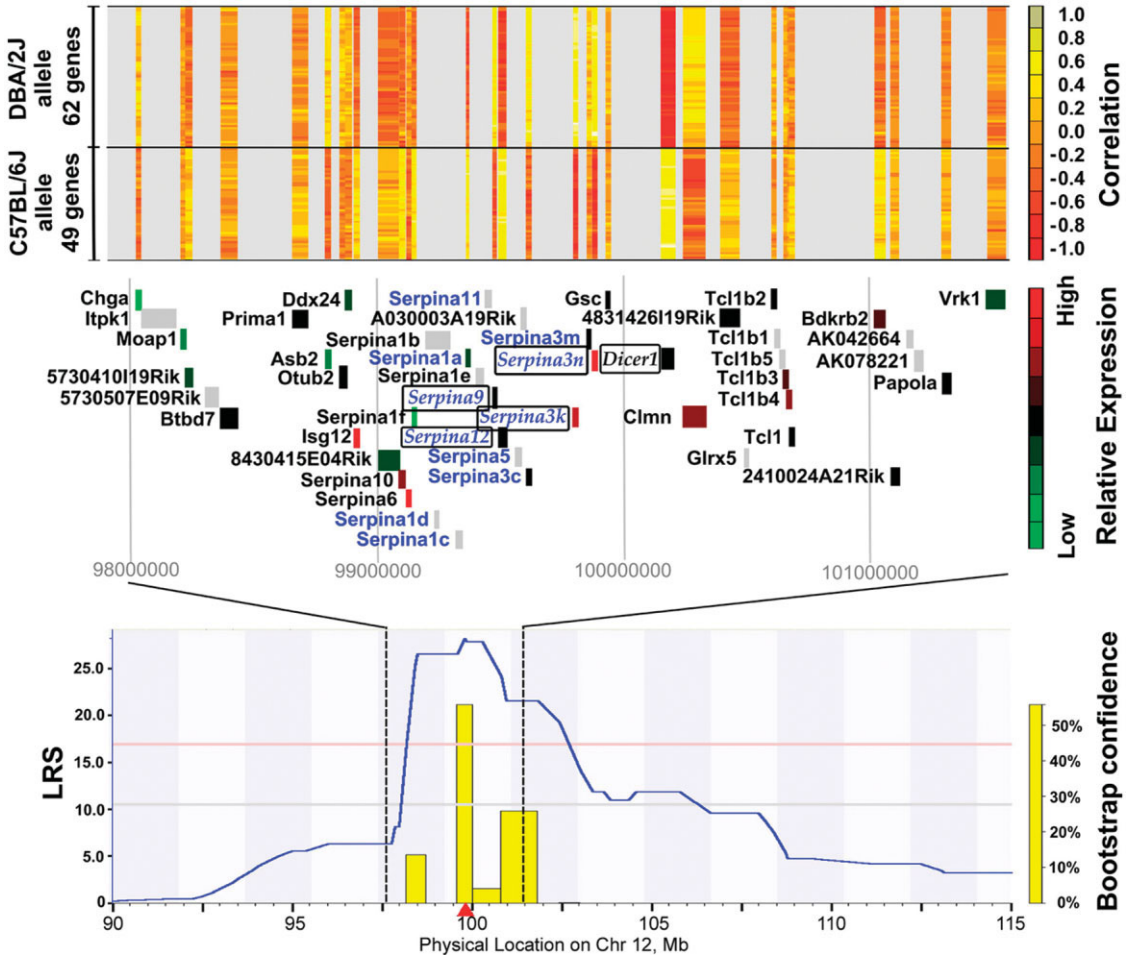

Fig. 4. The Chr12 locus is a master regulator of gene expression in mouse liver. A region on Chr12 that controls a large number of distantly regulated transcripts is shown in the bottom panel. The middle panel shows the genes in the QTL peak from 98-102 Mb. Genes labeled in blue text have nonsynonymous coding SNPs. Genes in italics are locally regulated. Genes in plain text are distantly regulated. The gene location marker is color-coded on the basis of the relative expression in the mouse liver. Green indicates low expression, red indicates high expression, and gray indicates that a gene was not represented on the microarray. The top panel shows the correlation between the expression of each gene located in this locus and the putative distantly regulated genes. Red indicates a negative correlation, yellow indicates a positive correlation, and gray indicates no data. Genes that are highly expressed in strains with either the DBA/2J or C57BL/6J allele at this locus are indicated.

around the 2 eQTL markers previously identified were considered (Fig. 4, lower panel).

The candidate gene that may be responsible for the variation in the expression between the transcripts associated with this locus should satisfy the following properties: (1) it should be locally regulated at the Chr12 locus or contain nonsynonymous coding SNPs between parental alleles, (2) it should have median to high expression in the liver, and (3) it should exhibit a strong correlation in the gene expression between the candidate gene and the distantly regulated transcripts when separated by a parental allele at this locus. The genes *Dicer1*, *Serpina3k*, *Serpina3n*, *Serpina9*, and *Serpina12* meet all 3 of these criteria and are likely candidates for the master-regulator gene in the liver.

Dicer1 is a logical candidate for the Chr12 locus because it is involved in posttranscriptional regulation of genes through the cleavage of double-stranded RNAs.[26] To date, no specific gene regulatory function has been proposed for Dicer1 in the liver. The expression was compared between the Chr12 locus–regulated genes in livers of wild-type and *Dicer1* heterozygous [the *Dicer1*-null genotype is embryonically lethal[27]] mice by quantitative reverse transcription PCR. As a negative control, a number of randomly selected genes that are not regulated by the Chr12 locus were selected. No consistent correlation was found between the expression of *Dicer1* and Chr12 locus–regulated genes (data not shown), and this suggests that *Dicer1* is not the master regulator at this locus. However, *Dicer1* heterozygous mice may not be the most appropriate system because *Dicer1* mRNA levels in heterozygotes are 134% of wild-type levels. Furthermore, a small sample size (n = 3) limits the power of the analysis ($P = 0.17$).

Other means of biological interpretation of the data were considered. GO and transcription factor binding site analyses were performed. A GOMiner[28] 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 ($P = 8.74 \times 10^{-4}$). The genes from this category that are distantly 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 distantly regulated transcripts at the Chr12 locus (LRS≥30) were divided into 2 groups: those with high expression when either the C57BL/6J or DBA/2J allele is present. The National Cancer Institute's promoter analysis tool 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[19] tool identified the Freac-2 site as significant in both C57BL/6J and DBA/2J lists ($P = 6.026e{-}02$ and $P = 1.823e{-}02$, respectively). Aryl hydrocarbon receptor nuclear translocator ($P = 8.577e{-}02$) and sex-determining region Y-box 9 (SOX-9) ($P = 9.159e{-}02$) sites were found to be common for DBA/2J allele-containing genes. The PAINT tool[20] was also applied to the data, and no significant transcription factor binding sites between the 2 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.[29] and Kulp and Jagalur[30] found that the genes in distantly regulated bands are not enriched for transcription factors or biological function. This suggests that the distantly 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, the murine orthologs of human serine protease inhibitor α1-anti-chymotrypsin. Although humans have 1 copy of α1-anti-chymotrypsin at 14q32, the mouse has 14 copies at 12E1.[31] 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 a function other than protease inhibition. Interestingly, human α1-anti-chymotrypsin has been reported to be able to bind to DNA and has been found to inhibit DNA polymerase and DNA primase in vitro.[32] Horvath et al.[33] 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.[34] However, it was also reported that human α1-anti-chymotrypsin binds to double-stranded DNA without specificity to known DNA binding motifs.[34] It has also been shown that some serpins may require cofactors for activation,[32] and this raises the possibility that the behavior of the *Serpina3* family changes, depending on the environment. Collectively, although there is no firm evidence for the role of *Serpina3* genes in the regulation of gene expression, we posit that our data point 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 RI and F2 crosses in both mice and rats along with the dense genetic marker maps for these strains. These data can be used to search for correlations between the phenotypes, gene expression, and genetic markers, that is, 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.

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*; Chr3, 3.620141 Mb), using WebQTL. *Hnf4g* is distantly regulated by the Chr12 locus (Fig. 5, left). Several of the BXD phenotypes are also regulated by this locus, and there is strong correlation (Fig. 5, right) between *Hnf4g* expression and the induction of serum interleukin-6 after tumor necrosis factor (TNF) injection, lethality due to TNF injection, and decreased body temperature after TNF injection.[35] Both the expression of *Hnf4g* and the values of these phenotypes are separated by the 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 knock-in mouse is neither the best option nor feasible. Thus, we propose that an 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 differ-

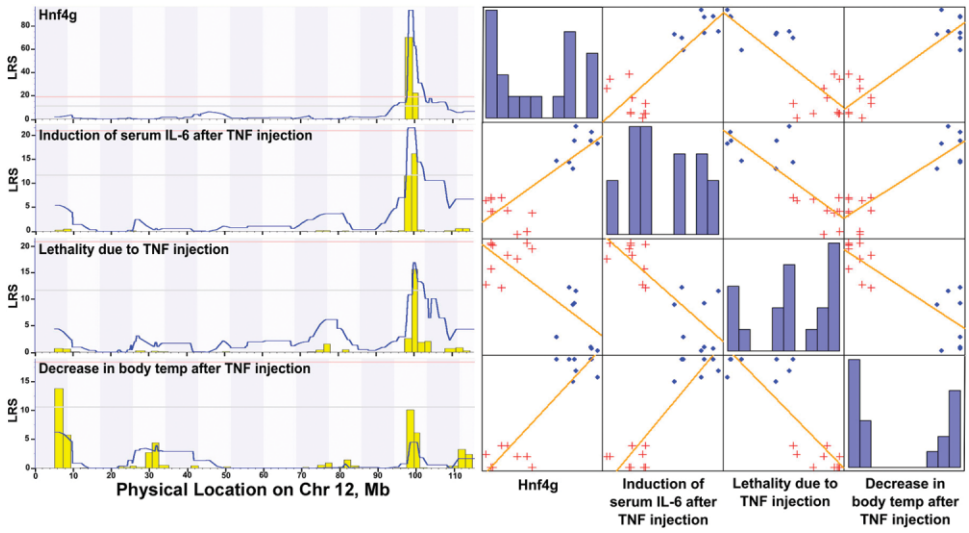

Fig. 5. In silico discovery of gene expression to phenotype correlations using WebQTL. The left panel shows QTL interval maps on Chr12 for hepatocyte nuclear factor 4, the gamma (*Hnf4g*) gene, and 3 related phenotypes. The right panel shows pairwise correlation plots for *Hnf4g* expression and the phenotypes. Each dot represents the measurements for 1 strain. Blue dots indicate strains in which the C57BL/6J allele is present at the Chr12 (99.83 Mb) locus. Red crosses indicate strains with the DBA/2J allele. Blue histograms along the diagonal represent the distribution of all values of the phenotype.

ences 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.[36,37] Accordingly, the liver expression data in WebQTL may be used to select strains that differ in the 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 monooxygenases are a class of phase 1 enzymes that oxidize organic nitrogen-containing and sulfur-containing compounds such as cimitidine, methimazole, and nicotine.[38-40] Flavin monooxygenase 3 (Fmo3) is the most common isoform in the human liver.[41] Although no sex differences in the human activity of Fmo3 have been shown, in mice, *Fmo3* expression has been shown to be much higher in females.[41,42] The basal expression of *Fmo3* varies widely across the BXD strains (Fig. 6, top). We confirmed this by quantitative real-time PCR analysis of *Fmo3* expression in 9 of 40 strains (Fig. 6, bottom). The correlation between the expression measured by the 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. Because *Fmo3* expression varies widely across BXD strains and mutations in this gene have been implicated in trimethylaminuria,[36] a disease in which trimethylamine is not metabolized but is excreted in the breath and sweat, 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 the liver. We describe several genetic loci that control the 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

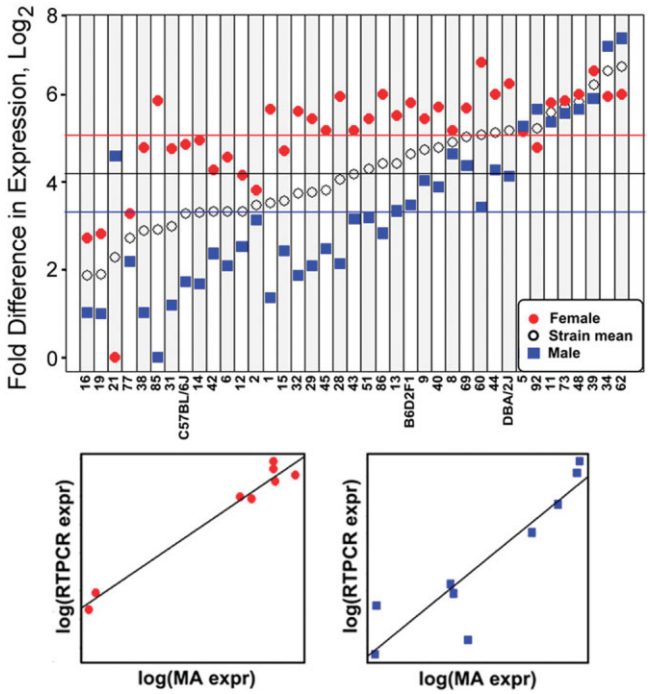

Fig. 6. WebQTL-assisted strain selection for phenotypic profiling: relative expression of *Fmo3* across BXD strains measured by a microarray. Red circles indicate values in females, blue squares indicate values in males, and black circles indicate strain means. The lower panels show relative *Fmo3* expression in select strains as measured by a microarray versus QRT-PCR. The expression of *Fmo3* in the BXD21 strain was set to 0.

liver and brain transcriptome maps, we highlighted tissue-specific differences in the regulation of gene expression. Finally, this study demonstrates how these publicly available data may be used to infer genotype-phenotype correlations, generate a testable hypothesis, and select mouse strains for further testing based on the genetically determined differences of expression of the key genes.

Acknowledgment: We thank Greg Hannon and Elizabeth Murchison from Cold Spring Harbor Laboratory for providing liver samples from *Dicer1* heterozygous animals.

References

1. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. Science 2002;296:752-755.
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.
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.
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.
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.
6. Taylor BA. Recombinant inbred strains. In: Lyon ML, Searle AG, eds. Genetic Variants and Strains of the Laboratory Mouse. Oxford, UK: Oxford University Press, 1989:773-796.
7. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 1989;121:185-199.
8. Farrall M. Quantitative genetic variation: a post-modern view. Hum Mol Genet 2004;13 Spec No 1:R1-R7.
9. 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.
10. Treadwell JA. Integrative strategies to identify candidate genes in rodent models of human alcoholism. Genome 2006;49:1-7.
11. 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.
12. 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.
13. Lee GH, Bennett LM, Carabeo RA, Drinkwater NR. Identification of hepatocarcinogen-resistance genes in DBA/2 mice. Genetics 1995;139: 387-395.
14. 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.
15. 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.
16. 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.
17. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics 1994;138:963-971.
18. Storey JD. A direct approach to false discovery rates. J R Stat Soc B 2002; 64:479-498.
19. 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.
20. 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.
21. Doss S, Schadt EE, Drake TA, Lusis AJ. Cis-acting expression quantitative trait loci in mice. Genome Res 2005;15:681-691.
22. Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet 2006;7:862-872.
23. 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.
24. 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.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289-300.
26. Hannon GJ. RNA interference. Nature 2002;418:244-251.
27. 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.
28. 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.
29. 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.
30. Kulp DC, Jagalur M. Causal inference of regulator-target pairs by gene mapping of expression phenotypes. BMC Genomics 2006;7:125.
31. 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.
32. Janciauskiene S. Conformational properties of serine proteinase inhibitors (serpins) confer multiple pathophysiological roles. Biochim Biophys Acta 2001;1535:221-235.
33. 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.
34. 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.
35. 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.
36. 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.
37. Rudnick DA, Perlmutter DH. Alpha-1-antitrypsin deficiency: a new paradigm for hepatocellular carcinoma in genetic liver disease. HEPATOLOGY 2005;42:514-521.
38. Ziegler DM. Recent studies on the structure and function of multisubstrate flavin-containing monooxygenases. Annu Rev Pharmacol Toxicol 1993;33:179-199.
39. 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.
40. 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.
41. 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.
42. 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.