Genome-Level Analysis of Genetic Regulation of Liver Gene Expression Networks

Daniel Gatti,* Akira Maki,† Elissa J. Chesler,* Roumyana Kirova,† Oksana Kosyk,‡ Lu Lu,§ Kenneth F. Manly,* Robert W. Williams,* Andy Perkins,* Michael A. Langston,* David W. Threadgill,*† and Ivan Rusyn*†

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. 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. 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. 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. 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.

The BXD panel of RI strains was created from the C57BL/6J and DBA/2J parental strains. At each locus,
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 brain3 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 na"ive 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 (qtreaper.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.abc.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
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.\textsuperscript{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 (\textit{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.\textsuperscript{22} This implies regulation by another gene, perhaps a transcription factor. The ADP-ribosyltransferase 5 (\textit{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 (\textit{Il21r}) gene, located on distal chromosome 7, is a good example (Fig. 1C). \textit{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 onChr12. (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., several 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. and the data from this study both show only a slight (54%-56%) enrichment for C57BL/6J high loci 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. 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. 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. 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 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 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\(^2\)). 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)
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. 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] 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 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: Bf3, Rqcd1, Gpr50, Tep10c, P2ry10, Olfr1403, Olfr1443, Olfr401, Olfr512, Olfr935, Olfr1341, Olfr341, Olfr656, Olfr1365, Mesp2, Met, Libp3, Fis3, 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 analysis tool found no common transcription factor binding present. The National Cancer Institute’s promoter analysis tool found no common transcription factor binding sites between the 2 lists that are distantly regulated by the Chr12 locus are mainly olfactory receptors: Bf3, Rqcd1, Gpr50, Tep10c, P2ry10, Olfr1403, Olfr1443, Olfr401, Olfr512, Olfr935, Olfr1341, Olfr341, Olfr656, Olfr1365, Mesp2, Met, Libp3, Fis3, Centd2, and Rassf3.

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. 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 knock-out 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-
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.\textsuperscript{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 cimetidine, methimazole, and nicotine.\textsuperscript{38-40} Flavin monooxygenase 3 (Fmo3) is the most common isoform in the human liver.\textsuperscript{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.\textsuperscript{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,\textsuperscript{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...
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