GENETIC REGULATION OF SEX-SPECIFIC GENE

EXPRESSION IN MOUSE LIVER

Ni Zhao

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Environmental Sciences and Engineering.

Chapel Hill

June 2009

Approved by:

Ivan Rusyn, MD, PhD

Wei Sun, PhD

Louise Ball, PhD

Abstract

Ni Zhao: Genetic regulation of sex-specific gene expression in mouse liver

(Under the direction of Ivan Rusyn, M.D., Ph.D.)

Sexual dimorphism in the expression of many genes is thought to play an important role in disease susceptibility, drug metabolism, and xenobiotic response in both humans and other species. While previous research has explored the relationship between phenotypes and sex-dependent differences in expression of individual genes, this study dissected the genetic underpinnings that control sex-specific gene expression in mouse liver. We performed genetic mapping of genome-wide liver mRNA expression data in naïve male and female mice from C57BL/6J, DBA/2J, B6D2F1, and 37 BXD strains. Thousands of liver transcripts exhibited considerable differences in expression between females and males. An array permutation based functional analysis identified several xenobiotic metabolism pathways, which are strongly dependent on subject's sex. Furthermore, expression quantitative trait locus (eQTL) mapping identified several eQTLs that are major sex-specific regulators of gene expression in mouse liver and the candidate genes that are likely to be the regulators for these loci were revealed. Co-expressed genes were shown to be more likely to be involved in similar functions, supporting the hypothesis of "guilt by association". *Conclusion*: This study provided more evidence in the sexually dimorphic gene expression in the liver, which can convey important implications to toxicological and pharmaceutical studies.

Acknowledgements

This work would not have been possible without the collaboration the following
individuals and institutions.   I would like to thank the mentoring of Dr Ivan Rusyn during
my master's program, the direction of Daniel Gatti, the help and guidance of our lab
manager Blair Bradford, and the support of every lab member in Dr Ivan Rusyn's lab.
Also I would like to thank the Department of Environmental Sciences and Engineering at
UNC for the opportunity to work and study in the United States.

Table of Contents

Page

List of Tables

List of Figures

List of Abbreviations

AIL - Advanced intercross line

BXD - Recombinant inbred cross between C57BL/6J and DBA/2J

FDR - False discovery rate

GO - Gene Ontology

LRS - Likelihood ratio statistic

QTL - Quantitative trait loci

RI - Recombinant inbred

CHAPTER 1

Literature Review

Sexual dimorphic phenotypes in the liver

Sexual dimorphism with respect to disease susceptibilities, drug metabolism and xenobiotic response are characteristics of many species, including humans. Liver is, arguably, the most metabolically active tissue and important for both pharmacology and toxicology and plays important role in defining the sex-specific differences. Metabolism of drugs and other chemicals in the liver may result in metabolic activation leading to either increased therapeutic effect (the statins, for example) or toxicity (acetaminophen, for example) in the liver or other organs. The physiological requirement for steroid hydroxylation differs between the sexes, and many enzymes in steroid metabolism, especially the cytochrome P450 superfamily, are expressed differently between males and females and this has long been known to have great biological implications (1; 2). For example, drugs metabolized by CYP3a family have long known to exhibit faster clearance in women than in men, although the clinical significance of this effect is yet to be confirmed (3). Furthermore, similar environmental exposure or diet can cause different effect to males and females because of their different metabolism in the liver. An anecdotal example includes alcohol induced liver injury, which is usually more severe in females than in males, with faster progression under similar alcohol consumption (4). The effect of the sexual dimorphic chemical metabolism is not limited to the liver, but influences other organs and the whole body as well. A mouse model has shown that carbon tetrachloride causes more severe

renal damage in male mice, an observation which was suggested to be associated with the reduction of the hepatic metabolizing capacity by CYP enzymes (5).

For pharmaceutical companies, regulators and the general public, drug safety is a serious concern and drug induced liver injury is the primary reason that a prospective drug is pulled off from clinical trials (6). Women and men have different susceptibility when receiving similar dose of drugs. It is reported that women experience more hepatic adverse effect to treatment with therapeutic drugs than men, with 74% of drug induced acute liver failure to be in women (7) In a 2002 report, women accounted for 79% of reactions due to acetaminophen and 73% of idiosyncratic drug reactions(8). However, traditional liver diagnostic tests are not able to identify women at risk of acute liver failure before drug administration, but identification of this problem after drug approval can lead to removal of drugs from the market (7). Moreover, if this sex differences happen when patients are intentionally exposed to drugs, similar sex differences should exist when men and women are exposed to environmental xenobiotics. Understanding how the liver metabolism differs between males and females can be important in understanding these sex-dimorphic physiologies.

With regard to baseline liver disease, women and men also show great difference in disease susceptibility, age of onset, or severity. Examples about sex dimorphic liver diseases abound in the literature. For example, chronic liver diseases, including chronic hepatitis B, C and hepatic steatosis, progress more rapidly in males than in females (9). Cirrhosis, which is an important pathological process in liver carcinogenesis, is mainly a disease of men and postmenopausal women (10). The incidence of hepatocellular carcinoma (HCC), the most common liver cancer, is higher in males than in females with a male:female ratio of 2:1 to 4:1

(11). The reasons for these sexually dimorphic phenotypes are still not completely understood and under constant exploration.

Sexual dimorphism through hormones

Many reasons have been proposed to be contributing to the hepatic sexual dimorphism. The most apparent and established reason is the sex hormone level that differs between males and females, beginning *in utero* and continuing throughout the life time of the organism (12; 13). Very few of the sex dimorphism can be apparent at birth, while most of the differences between sexes usually emerge at or after sexual maturation, when the sex hormone levels greatly differ between the two sexes. This indicates the role of sex hormones in regulating the sexual dimorphism. Two mechanisms have been proposed to explain the role of sex hormones in regulating sex specific phenotypes.

Firstly, sex hormones can regulate gene expression by directly binding to estrogen receptor (ER) or androgen receptor (AR). Both ER and AR belong to a family of nuclear receptors and works in similar mechanism in gene transcription regulation. In the absence of hormone, ER and AR are largely located in the cytosol. Hormone binding to the receptor causes the moving of the receptors to the nucleus, as well as configuration change and dimerization of the receptor. The dimerized receptors can bind to specific sequences of DNA known as hormone response elements and recruit other proteins which are responsible for the transcription of downstream DNA.(14; 15).The genes regulated by ER and AR include several essential molecules in key cellular processes, including immune responses, steroid metabolism, cell proliferation and apoptosis. The structure and function of ER and AR were best studied in cancers of reproductive tissues as well as in somatic tissues, such as liver (16-18).

The second mechanism by which the sex hormone affects gene expression is through sex specific secretion pattern of growth hormone (GH). GH is a pituitary protein hormone that regulates a broad range of physiological processes, including long bone growth, fatty acid oxidation, glucose uptake, and hepatic steroid and foreign compound metabolism. The sex specific secretion pattern of GH is most apparent in rat, where adult males secret GH in a pulsative manner, with about 2 hours low serum GH intervals while the females secret GH in a more frequent manner, resulting a constant blood level of GH. Similar patterns also exist for mice and humans, but to a smaller extend. These adult patterns of GH release are set during the neonatal period by exposure to gonadal steroids, which program the hypothalamus and its regulation of pituitary GH secretion at the onset of puberty and during adulthood (19). A major difference between male and female GH profiles is the GH free intervals between secretion pulses in males. The GH free interval is essential in the expression of a lot of male specific enzymes, including cytochrome P450 2C11 (CYP 2C11) (20). In the liver, the sex specific GH secretion activates intracellular signaling pathways to the sexually dimorphic transcription of CYPs and other liver-expressed genes.

Sexual dimorphism through gene regulation

The contribution of genes on sex chromosomes to human diseases and animal traits has long been appreciated. It is known for a long time that men and women have different sex chromosomes, with two X chromosomes in females and one X chromosome and one Y chromosome in males. The Y chromosome harbors very few genes, most of which are expressed only in testes. The other genes in Y chromosome are usually "house-keeping genes", which are essential in all tissues and conditions and expressed in a constant amount. These genes usually have X chromosome homologues that escape X inactivation (21). The X inactivation process

ensures that men and women share similar expression level for most of the X linked genes despite the different amount of DNA dosage. However, when there are genetic defect in some essential X chromosome genes, the X-inactivation pattern in females can be skewed in most cases in favor of the normal gene expression, which account for some of the sexually dimorphic diseases (22). For example, X-linked recessive diseases are mostly expressed in males: female carriers of X-linked recessive mutations often exhibit skewed X-inactivation pattern to produce normal gene expression and have healthy phenotypes (22). Anecdotal examples of these diseases include red-green color blindness, hemophilia, the Duchenne and Becker forms of muscular dystrophy (both of which involve mutations in the DMD gene) and the glycogen-storage disease. More generally, if this skewed X-inactivation can be responsible for the X-linked diseases, similar mechanism would be also possible contributors to other common diseases and phenotypes, in which X chromosomes genes are involved to a certain extent.

In contrast to sex chromosomes, the autosomal genomes are shared between males and females with common DNA sequences, gene structure and similar frequency of polymorphisms between males and females. The mechanism of how the autosomal genomes contribute to sex specific traits was not understood as clearly as the effect of sex chromosomes. The autosomal genomes were assumed to be similar between men and women until very recently. However, evidence is accumulating that natural variation within the autosomal genomes of many species also affect the physiological traits differing between males and females. Sex specific gene expression and regulation, other than DNA content, may underlie most phenotypic differences between males and females.

At the mRNA level, sexually dimorphic gene expression has been observed for many species in autosomal tissues, including flies (23), worms (24), fish (25), rodents (26) and

primates (27). For example, the sexual dimorphism in the regulation of oxidative stress response could potentially differentially affect the susceptibility of cardiovascular disease in males and females (19). A lot of these sexually dimorphic genes are essential in metabolism of steroid, drugs and other environmental chemicals. Examples include several cytochrome P450s, which are essential in drug and steroid metabolism (28-31). The observation of the sex specific gene expression pattern suggested that at least a part of the sexual dimorphism is attributable to the sex specific gene expression pattern. Some other studies have indicated that the sex specific gene expression is also tissue specific, whereby the genes differentially expressed in males and females in some tissues may not be sexually dimorphic in other tissues (32).

BXD Recombinant Inbred Mice

Mouse is the most used animal model in studying the effect of genetic background on gene expression. One benefit of using mouse models in genetic studies is that the genotypes of the mice can be controlled. Inbred strains are used in lab experiments for almost 100 years and its values in genetics are well established and appreciated. Inbred strains can be created by sibling mating for over 20 generations, leading to increased homozygosity in at least 99% of the genome. In this way, inbred strains are homozygous at all genomic loci and each strain has identical genotypes (33). When exposed to similar environmental toxicants or pharmaceuticals, mice of the same inbred strain would respond similarly. The decreased genetic diversity and phenotypic variance within an inbred strain can reduce the number of animals needed to detect statistical significance. Also the reduced phenotypic variation within each strain is invaluable for reproducing phenotypic measurement in different lab laboratories. Furthermore, a lot of the inbred strains have been sequenced or have extensive high density single nucleotide polymorphism (SNP) maps, which is essential in understanding the genetic regulation.

The human population is genetically and physiologically diverse: when people are exposed to similar environmental toxicants or other exposures, some of them may suffer no measurable injury while some people might experience severe damage. Each individual has different genotype, except for identical twins. Single inbred strains are not sufficient to understand the genetically diverse human population. Recombinant inbred (RI) strains are a special type of inbred strains, which are created in such a way that its genome is a permutation of the genomes of progenitor inbred mouse strains. BXD RI mice are created by mating two inbred strains C57B6J and DBA to get the F1 generation. Then the F1 generations are sibling mated to produce the F2 generation. The continuation of the sibling-mating process for more than 20 generations produced inbred lines in which the majority of the mouse genome is isogenic. Within each strain, every individual has identical autosomal chromosomes, while the panel of inbred mice can exhibit great differences in both genotypes and phenotypes. The C57B6J and DBA mice have been known to show great differences with regard to some important hepatic phenotypes. The BXD mice panel has been a useful tool in studies on certain disorders, including alcohol preference and tolerance (34), alcohol metabolism (35), responsiveness to aromatic hydrocarbons (36), N,N-diethylnitrosamine induced hepatocarcinogenesis (37) and diabetes and atherosclerosis (38; 39). The RI strains have a known ancestry and a controlled mixture of genomes while remain the homozygosity and reduced intrastrain phenotypic variance of inbred strains. It should be noted that, although individuals in each inbred strain have identical genome, different strains can exhibit extensive phenotypic differences and genetic differences, in accordance with genetically diverse human population. Studies on the RI strains can help understanding how the genetically diverse human population responds to different exposures.

Systematic evaluation of the sex specific gene expression in mice

The advance of gene expression technology has made possible to systematically evaluate the sex specific gene expression and explore how the natural variation within the autosomal genomes can affect the sex specific gene expression (40-42). Current microarray technology allows for the measurement of thousands of genes' expression simultaneously, with a reduced cost of time and money. The sequencing of entire mammalian genomes provided an unprecedented amount of genomic information, which can be combined with high throughput gene expression data to assess the effect of genetics on constitutive levels of gene expression(41-43). Quantitative Trait Locus (QTL) mapping can be used to associate a specific genotype with a phenotypic measurement like high density lipoproteins (44) and ethanol tolerance (45). When mRNA transcription, collected using microarray analysis, are considered to be quantitative phenotypes, similar QTL mapping strategy can be applied to uncover how the genetic background regulate the expression of genes.

From a whole genome prospective, five studies have investigated the sexually dimorphic gene expression in mouse liver (32; 46-49), two of which addressed the question of whether the GH regulate the sexually dimorphic gene expression (46; 47). A brief review of these studies will be helpful to our understanding of the genetic reasons for sexual dimorphism in mouse liver.

Amador-Noguez et al (47) used gene expression profiling to determine the sex-specific and sex-independent changes in the liver of Ames dwarf mice compared with their wild-type litter mates. Ames dwarf mice are homozygous with a loss of function mutation in the *Prop1* gene, which can cause defect embryonic development of anterior pituitary gland and lead to life-long deficiencies in GH, prolactin, insulin like growth factor, and thyroid stimulating hormone secretion (50; 51). Out of the 14,000 genes mapped to the microarray, 123 were discovered as

significantly differently expressed between two sexes in wild-type mice with p values less than 0.001. In contrast, the sexually dimorphic gene expression pattern is nearly completely lost in the Ames dwarf mice, with only 7 genes differently expressed in two sexes. The fact that *Prop1* mutation produced a nearly complete loss of sexual dimorphism in gene expression indicates that pituitary gland hormones are a major contributor to the sex specific gene expression. In a study by Clodfelter et al (52), large scale gene expression profiling was used to evaluate the expression in wild type and Stat5b inactivated mice and(53) to characterize sex differences in liver gene expression and their dependence on Stat5b. STAT (signal transducer and activator of transcription) proteins are transcription factors which respond to a variety of extracellular cytokine and growth factor signals, especially growth hormone (54-56). At a fold change threshold of 1.5, 1603 genes were identified with sex biased expression in wild type mice. Of the 850 genes which showed higher expression in males, 90% were down-regulated in Stat5b deficient males. Similarly, out of the 753 genes with higher expression in females, 61% was upregulated in Stat5b deficient males. However, 90% of the sex biased genes in females were not affected by Stat5b deficiency. These findings suggested that Stat5b is essential in the expression of male-predominant genes. A following study (46) by the same group investigated the dependence of sex specific liver gene expression on STAT5a. STAT5a is another isoform of STAT5, which shares 90% identical amino acid sequence to STAT5b and also responds to sexually dimorphic plasma GH stimulation. 1437 genes were discovered as female biased in wild type mice, within which 219 (15%) had decreased expression in STAT5a-deficient females. In contrast, in the 1045 male biased genes, only 56 had significantly increased expression in Stat5a knock out females, the effect of which is trivial. These two studies, together, provided evidence of the importance of STAT on the regulation of sex-specific gene expression, with STAT5a more

16

significant in female livers while STAT5b more important in male livers. Because both STAT5a and STAT5b are transcription factors involved in GH signaling pathway, these results confirmed the important role of GH in regulating the sexually dimorphic gene expression.

The other two studies (32; 49) differ from the previous three by, instead of assessing the sexually dimorphic genes in a single strain, evaluating the gene expression differences in several strains. Yang et al (32) analyzed the gene expression from 334 mice derived from an intercross between inbred mouse strains C57BL/6J and C3H/HeJ in multiple tissues, including the liver. All strains used were apolipoprotein E (apoE) knockout mice that were sacrificed at 6 months of age after being fed a Western diet for 4 months. 9250 (72%) genes were detected as sexually dimorphic in the liver, indicating the widespread sexual dimorphism in hepatic gene expression. But when the extent of sexually dimorphism was considered, most of these sexually dimorphic genes displayed <1.2 fold change in one sex versus the other. These sexually dimorphic genes in the liver are enriched for protease inhibitor activity, immune/defense response, carboxylic acid, fatty acid, steroid and lipid metabolic pathways, electron transport, monooxygenase activity, and oxidoreductase activity. Moreover, the eQTL hotspots for subsets of the sexually dimorphic genes provided evidence of the genetic regulation of gene expression. The use of the apoE knockout mice and the high fat diet was related to a sex biasd trait---atherosclerosis. Although the authors could not rule out the influence of the particular genetic background or the effect of diet on sexually dimorphic gene expression, the large number of animals provided the power to detect the small differences in gene expression between sexes and the intercross mouse genome allowed the analysis of genetic control of sexually dimorphic gene expression.

The very recent study, Su et al used a customer designed microarray to evaluate the sex and strain effect on gene expression and exon expression in the liver of three mouse strains -

DBA/2J, C57BL/6J and C3H/HeJ. Gene expression was assessed using two methods: 3' gene expression profiling and whole-transcript gene expression profiling. Exon expression was determined using exon probes and flanking junction probes that spanned across the neighboring exons. The exon expression was reflective of not only the mRNA transcription, but also alternative splicing. 32% of the genes have sex biased expression in at least one of its exons while only 17% of the genes were detected as sex biased using 3' gene expression profiling (Bonferroni corrected p-value < 0.01). Over 90% of the genes that were identified as sexually dimorphic via either whole gene expression profiling were also identified as sexually dimorphic via exon profiling. On the other hand, 38% of the genes identified as sexually dimorphic using exon profiling cannot be identified via either 3' gene profiling or whole gene profiling. The fact that exon profiling identifies more differences indicated that sex also influence the alternative splicing process as well as gene transcription, providing suggestions for further studies.

Missing from the current literature is a study of constitutive gene expression in the sexually dimorphic gene expression in mouse liver using a panel of RI strains. This kind of study would be revealing to why men and women respond differently to similar environmental and pharmaceutical exposures. Also it would help us understand how small changes in a cluster of functionally related genes can have significant effect on physiology.

CHAPTER 2

Introduction

Despite having nearly identical genomes, males and females have been shown to differ in gene expression, disease susceptibility, drug metabolism, and xenobiotic response (4; 57). The liver is a key organ for defining sex-specific differences in steroid and xenobiotic metabolism, as well as complex physiology and function of other tissues (58). Much is known about the genetic and transcriptional regulatory mechanisms that control sex-specific differences in response to certain xenobiotics in the liver (32; 59-61). Evidence is sparse, however, regarding the extent of the global differences in gene expression networks that may exist between males and females under normal physiological conditions (62).

Gene expression quantitative trait locus (eQTL) mapping is a statistical approach used for comparing mRNA levels, collected using microarray analysis, with genetic polymorphisms segregating in a population to discover genomic intervals that are likely to regulate the expression of each gene (63). This approach has been used successfully to identify co-regulated genes, to discover genes potentially regulating the expression networks, and to understand normal tissue-specific physiology and underlying disease-related phenotypes (40-42; 64; 65). Successful validation of a *trans*-eQTL hotspot relating to oxidative phosphorylation in mouse adipose tissue was recently reported (66), thus underscoring the validity of this computational approach for identification of key networks that may be genetically controlled.

With regards to the liver, the sex-independent polymorphic local and distant QTLs, including several loci that control the expression of large numbers of genes, were also identified

by comparing the physical transcript position with the location of the controlling QTL (67). Still, it is not known whether sexual dimorphism exists in genetic regulation of liver gene expression networks. To address this gap in our knowledge, we used gene expression data from livers of naive mice from C57BL/6J, DBA/2J, B6D2F1, and 37 BXD strains to understand the variation in gene expression between males and females. Genes that were differently expressed between sexes were selected and assessed. Pathway analysis was carried out using an array permutation approach to uncover the biological pathways exhibiting strong sexual dimorphism. eQTL mapping and transcriptome maps of both male and female liver gene expression were compared to discover the similarities and differences in regulatory networks.

CHAPTER 3

Materials and Methods

Gene Expression Data

The details on mouse breeding, housing, RNA isolation and gene expression are described elsewhere (67). Tissue collection was conducted at the University of Tennessee at Memphis and approved by the Institutional Animal Care and Use Committee at UT-Memphis. Briefly, 37 strains of male BXD RI mice, C57BL/6J and DBA/2J parentals, and B6D2F1 were used to perform genome-wide eQTL mapping for 20,868 transcripts using Agilent (Santa Clara, CA) G4121A microarrays. RNA was pooled from 2-3 mice of the same sex and strain for each microarray and each strain has only one microarray for each sex. Gene expression data is available from www.genenetwork.org.

Statistical Analysis of Gene Expression Data

A Student's t-test with a Bonferroni correction ($\alpha = 0.01$) was used to select genes that are differentially expressed between females and males. These genes were grouped by fold-change (>1.2, >1.5, >2, and >3) differences between sexes using the median $\log_2$ transformed expression value in each sex. A two-step permutation based method (68) Significance Analysis of Function and Expression (SAFE) was used to identify the significant functional relationships among differentially expressed genes using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) using a graphical interface SAFEGui v1.0 (69). SAFE settings were as follows: the Student's t-test was chosen as local statistic, the Wilcoxon Rank Sum test as global statistics. Benjamini and Hochberg false-discovery rate (FDR) was determined under

5000 permutations. The graphical algorithm of Voy et al (70) was utilized to assess gene correlation. A gene expression "clique" was defined as a cluster of genes in which every gene pair has an absolute Pearson correlation coefficient above 0.9.

eQTL Mapping

QTL linkage mapping was carried out using QTLReaper (71) which performed 1000 permutations of the strain labels to obtain genome wide p-values (72). Interval mapping was performed in a web-based tool WebQTL (73). The details of the markers, QTL identification and WebQTL were reported elsewhere (67). Microarray probes were mapped to the mouse genome (NCBI Build 36) and SNPs overlapping the probes were sought using data from Szatkiewicz et al. (74). The proportion of SNPs overlapping probes for cis-QTLs was compared with that in trans-QTLs using a χ^2 test to determine the potential bias. Transcriptome maps for both sexes were produced in R [ver. 2.7, (75)] by relating transcripts location to their maximum LRS location on the genome at FDR<=0.25, using the QTL data from QTLReaper. Histograms were generated by counting the number of transcripts mapped to each genomic location within 1 Mb window. Genomic regions that are identical by descent (IBD) between C57BL/6J and DBA/2J were inferred by searching for 100 consecutive SNPs with identical allele calls in both parental strains using the web tool [http://compgen.unc.edu/DisplayIntervals/DisplayIntervals.html; (76)]. The IBD regions between C57BL/6J and DBA/2J were excluded from the analysis to narrow eQTL windows.

Transcription Factor Binding Site (TFBS) Enrichment

The oPPOSUM (77) tool was used to assess the TFBS enrichment in the differentially expressed genes of each sex for Single Site Analysis (78). For each transcript, the top 10% of conserved region in the 5000 bp upstream/downstream sequences between mouse and humans

with minimum conservation of 70% and matrix match threshold of 80% was scanned for TFBS

in the JASPAR database using a position weight matrices algorithm.

CHAPTER 4

Results

Characterization of sexually dimorphic liver genes in BXD mouse panel

Genes that are differentially expressed in female and male livers were selected using a student's t-test and a Bonferroni correction ($\alpha = 0.01$). Out of the 20,868 transcripts on the array, 1,534 genes were found to be sexually dimorphic. Specifically, 639 genes had significantly higher expression in females (female-biased genes) and 895 in males (male-biased genes). Table 1 summarizes the distribution of sexually dimorphic genes, at different fold-change level. The majority of sexually dimorphic genes (72.6% in females and 84.9% in males) are less than 1.5-fold different between sexes.

To determine whether the differences could be attributed to control by transcription factor(s), binding site enrichment was evaluated in the conserved sequences within 5,000-bp upstream and downstream regions of the sexually dimorphic genes. The sexually dimorphic genes (1,534 genes) showed significant enrichment for the binding site for HNF4A. The female-biased genes (639 genes) showed enrichment for the binding site for HNF1A, IRF2 and NR2F1; however, the male-biased genes (895 genes) didn't exhibit evidence for transcription factor binding. Androgen binding site was not discovered as significant TFBS in the sex biased genes (Fishers exact p = 0.379) and estrogen binding site is not quarried by oPOSSUM tool.

SAFE (68) software was used to determine biological categories and pathways associated with sexually dimorphism in gene expression. Using an array permutation approach, SAFE takes into account unknown correlations between genes in functional categories. After accounting for

multiple comparisons, none of the Gene Ontology pathways were shown to be significant at false discovery rate 0.25. Nine KEGG pathways were significantly different in males and females at FDR 0.05 (Table 2) and 59 KEGG pathways were significant at FDR 0.25.

Genetic regulation of sexually dimorphic genes in mouse liver

To investigate possible genetic control of sex differences in gene expression regulatory networks, interval mapping was independently performed in each sex. A transcript which has a QTL less than 5 Mb away from its own genomic location was considered to be a cis-QTL (79; 80), implying a possible controlling mechanism near the gene itself, although the mere existence of cis-QTL is not sufficient to draw firm conclusions on the regulation mechanism. QTLs which are more than 5 Mb away from their controlled transcripts were considered as trans-QTLs. When a transcript showed the same maximum QTL location in both sexes, this QTL was identified as a shared QTL. It has been suggested that, since C57BL/6J sequence is used for array probe design, spurious eQTLs may be detected in genetic crosses due to SNPs in probes (79; 81). We found that no bias exists in our data (data not shown).

The distribution of eQTL in both sexes is summarized at Table 3. It is interesting that cis-eQTLs are more robust than trans-eQTLs whereby more cis-eQTLs were shared between females and males. In addition, more cis-eQTLs remained significant with increased stringency (FDR threshold) while the number of trans-QTLs diminished rapidly.

In both female and male datasets there are genomic loci that appear to regulate a larger number of genes than expected by chance. The, so called, eQTL hotspots are SNPs that are associated with expression differences in ten or more transcripts (Table 4). The cutoff was selected based on the 95 percentile of the binomial distribution with n=20,868 (the number of transcripts analyzed) and probability = 1/3,795 (the number of SNP markers).

Sex-specific transcriptome maps for liver gene expression in mice were generated by plotting location of each transcript with an eQTL against the location of its respective peak eQTL location (Figures 1A and B). Histograms were generated by counting the number of transcripts mapped to each genomic locus within 1 Mb window (Figures 1 C and D).

In addition to sex-independent strong "master-regulator" trans-eQTL that has been reported to exist in mouse liver (67; 82), there are several sex-specific eQTLs that control expression of dozens of liver genes (Figures 1C and D). To identify the potential regulatory genes in the QTL transbands, we first narrowed the QTL regions by eliminating regions that are identical-by-descent between C57BL/6J and DBA/2J mice. High density SNP maps encompassing 7.87 million polymorphic loci across these strains show that about 43.7% of the genome is shared between C57BL/6J and DBA/2J strains, which is in agreement with the breeding history (83). Regions that are not identical-by-descent were further explored for presence of sex-specific "master-regulator" genes (Figure 2A and B). Candidate genes were selected using the following criteria: 1) it should be locally regulated or contain non-synonymous coding SNPs between C57BL/6J and DBA/2J mice; 2) it is expressed in the liver. *Naps1, Ceacam9* and *Grlf1* are candidate regulatory genes that may be responsible for the male-specific chromosome 7 eQTL hotspot (Figure 2B); however, no annotated genes were identified in the strongest female QTL hotspot on chromosome 6 (Figure 2A).

Gene Correlation exhibits distinct patterns from gene expression between the sexes.

Co-regulated genes are more likely to be co-expressed and correlation matrix network analysis has been proposed for discovery of genes that form "cliques" (70). We applied this approach to liver gene expression data from male and female mice separately. Gene pairs with absolute Pearson correlation coefficient greater than 0.90 were selected to construct edge-

weighted graphs of the cliques. The top 30 genes appearing in the largest number of cliques were defined as clique enriched genes. There was little overlap between clique-enriched genes and genes that are differentially expressed between sexes (data not shown). For example, only nine of the top 30 male clique-enriched genes also exhibited higher expression in males, while none of the top 30 female clique-enriched genes had significantly higher expression in females.

To visualize clique-enriched genes, a correlation matrix (Figure 3A and D) and a gene expression heat map (Figures 3B and E) were constructed using unsupervised two dimensional clustering analysis of the 579 female and 625 male transcripts which are involved in at least one clique. Interestingly, highly correlating genes show distinct patterns in expression between strains (red and blue bars for sub-clusters) whereby some clusters (blue bars) are clearly sex-specific, while others (red bars) are not. With regards to sex-specific co-correlating genes, there were 76 in females and 151 in males and 60 of those were shared between these clusters.

In addition, when eQTL locations for all co-correlating genes were plotted (Figures 3C and F) it became evident that highly correlated transcripts, represented by cliques, are more likely to be regulated by similar eQTL loci. For example, 75 (49.6%) of the 151 male-specific co-correlated transcripts, which varied in expression between sexes (blue bars), share an eQTL at chromosome 5 (58.9 Mb to 71.0 Mb) and are highly inter-connected in a "molecular transport" ingenuity-derived network (Figure 4B). Similarly, 49 (64.5%) of the 76 females-specific genes have their eQTLs clustered on distal chromosome 5 (136.5 Mb to 144. 7 Mb). The top network for the female genes was "cell-cell signaling and interaction" (Figure 4A).

CHAPTER 5

Discussion

Sex-dependent differences in gene expression in the liver: physiological and toxicological significance

In this study, we studied the magnitude and possible regulation of sex-dependent differences in gene expression in liver using a panel of BXD recombinant strains. While several reports have examined sexual dimorphism in liver gene expression in the mouse, our work has used a genetically-defined population (84) to address this issue on a more comprehensive level and to determine whether potential genetic regulators can be identified.

Indeed, thousands of genes were identified as significantly differentially expressed between males and females in the 41 strains at different fold change levels with many of the previously reported genes also confirmed by our study. A recent review article (62) provides a summary of four earlier studies in mouse liver (32; 46-48) and identifies 48 genes as being at least two-fold higher expressed in females than in males in at least two studies. Of these, 26 genes show a comparable magnitude of female-biased expression in our study (Table 5A). Similarly, 15 of 54 males-biased genes selected using the same criteria were confirmed in our work (Table 5B). Since previous studies were conducted using varying experimental designs (different strains, diet, age, and physiological conditions), a wide variety of microarray platforms and statistical stringency criteria, we consider this high (28 and 54%) degree of replication to be

biologically significant as the genes that replicate between multiple studies are highly robust representatives of sex-specific expression differences.

Two previous studies (32; 49) assessed sex-specific differences in liver gene expression using more than one strains of mice and reported a list of sexually dimorphic genes at comparable threshold cutoffs. The magnitude of sexual dimorphism in liver transcriptome (7.4% transcripts showing the effect of sex) in our study is smaller than reported previously. For example, Su and co-workers (49) reported that about 17% genes are sexually dimorphic in three mouse strains – DBA/2J, C57BL/6J and C3H/HeJ. Yang et al (32) concluded that as many as 72.0% (9250 of 12,845) of the genes are sex-biased in an F2 cross between C3H/HeJ and C57BL/6J. While the exact reason for the differences between our and other studies is difficult to determine, both statistical methods and study design factors are most likely contributors. First, Su et al (49) reasoned that a very high percentage of sexually dimorphic genes in their study could be due to the number of animals assessed per sex per strain in different studies. Second, when our data was re-analyzed using the criteria detailed by Yang et al (32) (i.e., Q value=0.2%), a larger proportion of genes (17.2%) 3597 genes was found to be sexually dimorphic, a finding confirming the use of Bonferroni correction in our work to establish a more robust representation of sexual dimorphism in gene expression.

The sexually dimorphic expression of many drug metabolism and other liver-expressed genes has been suggested to be regulated, to a large degree, by the temporal pattern of plasma growth hormone release by the pituitary gland, which shows significant sex differences. These differences are most pronounced in rodents, where plasma growth hormone profiles are highly pulsatile in males but are nearly continuous in females. The growth hormone–Stat5b pathway was shown to be one of the key regulators of sexual dimorphism in mouse liver gene expression.

Hepatocyte nuclear factor 4a (HNF4a) is an important transcription factor in response to sexually dimorphic growth hormone secretion pattern responsible for up-regulation of males-biased genes (60; 85). Indeed, our study showed that HNF4a binding site was significantly enriched in sexually dimorphic genes, consistent with the biological roles of HNF4a and growth hormone.

Consistent with previous reports, the reproducible sexually dimorphic genes detected in this study are essential for lipid metabolism, xenobiotic response and hormone metabolism (Table 2). For example, *Gstp1* is a male-biased gene which encodes a key glutathione S-transferase enzyme. In humans, GSTpi plays an important role in the detoxification of many hydrophobic and electrophilic compounds and is a biomarker of the overall survival in cancer patients (86). In the mouse, even though the baseline expression of *Gstp1* is much higher in male liver, its expression is suppressed in males, but elevated in females in pre-neoplastic liver lesions (87). Flavin-containing monooxygenase 3 (*Fmo3*), on the other hand, is a prominent female-biased gene in the mouse liver, but in humans *Fmo3* is expressed in livers of both sexes (88; 89). *Fmo3* has affinity for numerous substrates, including nicotine, tertiary amines, drugs, carbamates and organophosphates (90). This suggests that sex differences in *Fmo3* expression should be acknowledged in mouse toxicity studies in which *Fmo3* is the enzyme responsible for metabolism of the xenobiotic under investigation.

Genetic component of the sex specific gene regulation networks: eQTL analysis

It is likely that sex-specific hepatic gene expression has a complex mode of regulation with mechanisms other than growth and sex hormones being involved. The sex-specific eQTL hotspots discovered in our study provides evidence for the genetic component that may be also important. Previously, Yang *et al.* (32) assessed whether genetic variation can regulate sexually dimorphic gene expression by using eQTL analysis. We utilized a different strategy by searching

for QTLs in males and females separately. Four loci were identified to be regulatory hotspots in both males and females, including the strongest regulatory loci in chromosome 12. Six loci were found to be regulatory in females specifically and another five distinct loci were discovered in males.

We report that a locus on chromosome 7 is a most significant male-specific eQTL hotspot. *Naps1, Ceacam9* and *Grlf1,* are located in this location and are possible quantitative trait genes. *Npas1* (neuronal PAS domain protein 1) encodes a basic helix-loop-helix transcription factor which inhibits gene transcription (91), and is a reasonable candidate for regulation of gene expression. In adult mice *Npas1* is only expressed in specific regions of the brain and Npas1 negatively regulates the expression of erythropoietin and promotes neuronal progenitors in the nervous system (92). *In vivo*, *Npas1*-knockout mice exhibit behavior deficiency, including diminished startle response, as measured by prepulse inhibition, and impaired social recognition (93). In mouse embryos, *Npas1* has been found in a variety of tissues including liver and the Npas1 protein was suggested to be stable and abundant in the liver, despite relatively low mRNA levels (94). The possibility exists that the embryonic liver gene expression causes long term effects in regulating the expression of other genes. It is also possible that the Npas1 protein persists in the liver, as a transcription factor through adulthood. In humans, *NPAS1* was mapped to chromosome 19q13.2-q13.3, a syntenic region to the region in mouse chromosome 7 containing *Npas1*, indicating that the mouse and human genes are true homologs (95). *Ceacam9* (carcinoembryonic antigen-related cell adhesion molecule 9) is a member of a family of glycoproteins containing immunoglobulin domains (96). . However, its function is poorly understood and knockout mice developed no observed abnormalities during development. *Grlf1* (glucocorticoid receptor DNA binding factor 1, p190A) encodes a Ras GAP-binding

phosphoprotein which regulates the actin cytoskeleton and has important function in cell adhesion, migration and polarity (97). The encoded protein can associate with the promoter region of the glucocorticoid receptor gene and suppress glucocorticoid receptor transcription. In human cancer cell lines, Grlf1 expression is regulated by glucocorticoids may act as a human tumor repressor gene (98).

While the strongest eQTL liver locus in both sexes is on chromosome 12 (82), the strongest male-specific eQTL transband is on chromosome 7 and regulates the expression of 55 genes. The strongest female-specific eQTL transband is on chromosome 6 and it regulates only about half that many transcripts (26 genes). It is possible that a much weaker female bias in genetic regulation of gene expression is due to the lack of synchronization of the estrous cycle in our study; thus, a study assessing the influence of estrus cycle on the global gene expression or one using synchronized females may minimize this potential bias.

Co-correlation analysis reveals sex-specific gene expression networks

The correlation analysis revealed that gene co-expression networks may be used as a tool to uncover the connectivity between gene function and gene regulation. The fact that highly correlated genes are involved in similar functions supported the assumption of "guilt-by-association" (99). The graphical algorithm (70) used here allows for the assignment of each gene into different connected clusters, consistent with the biological context whereby genes may play important roles in multiple distinct pathways. Also, unlike other methods, this analysis doesn't seem to be plagued with false positives because of the stringent criteria to form cliques. However several limitations do exist in this analysis: because cliques represent the perfectly correlated gene clusters and the selection of cliques relies on an edge meeting a high threshold (correlation

coefficient of 0.9), it can exclude the edges which fall short of the selection threshold. This might impede biologically interpretation of the gene-gene relationships.

In conclusion, this study assessed the hepatic gene expression network between two sexes in a BXD RI mice population. Unlike studies carried in only one strain, the genetic diversity in the BXD population would shed light on the global sexually dimorphic gene expression in a human population with similar genetic and phenotypic diversity. The well studied inbred mouse genome provides gene and function annotations which allow us to generate testable hypothesis with regard to gene regulation network. Also our findings have several implications for toxicology and pharmaceutical studies. Not only the few genes that show significant sex differences are important in assessing toxic responses in males and females, a large number of functionally related genes with small differences between sexes could also contribute to many sex biased phenotypes. Cautions should be taken when applying the results from a study which were carried out in only male objects (which is very common in toxicological studies) to the whole population. Finally, the sexually dimorphic gene expression, at least partly, is due to the genetic reasons. Other genomic information, like genome structure similarities and IBD regions, when combined with eQTL mapping, can facilitate the process in finding candidate regulatory genes.

CHAPTER 6

Study Limitations and Further Direction

This study does have some limitations and further research would be needed to better elucidate the mechanism of sexually dimorphic gene expression pattern and how these results can be applied in human risk assessment.

First of all, the way the data was generated in this current study doesn't allow for the assessment of sex-strain interaction with regard to gene expression regulation. In this study, RNA from 2-3 mice of the same sex and strain was pooled in a single microarray plate to get the gene expression. This pooling procedure has the benefit in smoothing the gene expression variation in different individuals within the same strain and eliminating possible outliers. However, only one microarray data point was generated per sex per strain, therefore the effect of sex and strain interaction on gene expression cannot be analyzed. The study can only provide preliminary insight and testable hypotheses about the genes responsible for gene expression regulation. On the other hand, one recent study conducted on DBA, C57B/6J and C3H/HeJ mouse strains has indicated that the sex differences are larger than strain differences in gene expression (49), suggesting the importance of testing the sexually dimorphic gene expression in understanding liver biology even without the analysis of sex-strain interaction.

Secondly, the regulation of sex specific gene expression is very complex, with a lot of mechanisms contributing. The result of the genetic regulation of gene expression cannot be conclusive because of the inability to distinguish the effect of sex hormone and growth hormone on the sex-specific gene expression. Several methods are possible solutions. One alternative

option is to culture hepatic cells *in vitro* and assess the constitutive gene expression. After culturing in hormone-devoid medium environment, the effect of sex hormone on RNA level would be eliminated or reduced, leaving the genetic reasons for sex specific gene expression to be detected. Mouse with ovaries or testes removed would be an *in vivo* model to eliminate the sex hormone effect on gene expression. The organ removed animals can create similar hormone-devoid inner environment, leaving the genetic reasons regulating sex specific gene expression to be controlled.

Up to now, no study has assessed how the global gene expression differs in females in different estrous cycle status in any somatic tissue, nor did any study assess the gene expression on synchronized females. It is known that the estrous cycle in females can influence the expression of many genes in the liver, especially the genes that are involved in steroid hormone metabolism (100; 101). The discovery that the female eQTLs appear to regulate the expression of fewer genes are indications that female gene expression regulation may be more complexly regulated, with estrous cycle and hormone fluctuation being possible reasons. A further analysis on gene expression in different estrous cycle would further our understanding of genetic regulation.

In summary, this study provided more evidence on the sexually dimorphic genetic regulation of gene expression in the mouse liver. Studies on mouse models could be the foundation of our understanding of genetically diverse human population, in which huge differences exist between men and women with regard to disease susceptibility, drug metabolism and xenobiotic response. Although not conclusive, this study pointed to the sex specific gene expression regulatory network in mice. Further studies will be needed for further elucidation of the effect of sex as contributors to liver biology.

Table 1. Distribution of genes that are differently expressed in males and females, separated by fold change (p<0.01, Bonferroni correction)

Fold change	Dimorphic genes	Female high	Male high
>1	1534	639	895
>1.2	1394	562	832
>1.5	549	175	374
>2.0	183	48	135
>3	24	12	10

Table 2: The significant sexually dimorphic KEGG pathways discovered by an array permutation based method "SAFE".

KEGG	Size	Empirical p-value	Adjusted p-value	Description
KEGG:00630	14	0	0.017	Glyoxylate and dicarboxylate metabolism
KEGG:00830	54	0	0.017	Retinol metabolism
KEGG:00982	66	0	0.022	Drug metabolism - cytochrome P450
KEGG:00150	37	0.001	0.034	Androgen and estrogen metabolism
KEGG:00590	66	0.001	0.041	Arachidonic acid metabolism
KEGG:00480	40	0.002	0.045	Glutathione metabolism
KEGG: 00564	59	0.003	0.048	Glycerophospholipid metabolism
KEGG:04512	89	0.002	0.049	ECM-receptor interaction
KEGG:00980	58	0.004	0.049	Metabolism of xenobiotics by cytochrome P450

Table 3: Distribution of QTLs for each sex at different FDR level.

FDR		Cis-QTL	Trans-QTL	Total	QTL in both sexes
	Female	815	855	1670	
0.25	Male	719	780	1499	2465
	Shared	533	111	644	
	Female	632	276	908	
0.05	Male	547	274	821	1402
	Shared	414	81	495	
	Female	493	165	658	
0.01	Male	430	179	609	1091
	Shared	325	64	389	

Table 4: The SNPs which show the highest LOD score for more than 10 transcripts

(* indicates same SNP marker in males and females)

Females				Males			
SNPs	Chr	Genomic location (Mb)	Number of transcripts	SNPs	Chr	Genomic location (Mb)	Number of transcripts
rs8256197*	1	130.4166	22	rs8256197*	1	130.4166	15
UT_2_119.151187*	2	315.6424	13	UT_2_119.151187*	2	315.6424	16
rs13477796	4	617.4422	12	rs13479126	7	1006.577	24
rs6258088	4	621.0064	16	rs3675839	7	1010.672	11
rs6404906	4	672.1987	14	rs6295100	7	1011.097	58
rs13478831	6	922.0632	28	rs4226520	7	1024.968	11
rs13481087	11	1596.074	11	rs13479813	8	1211.448	13
rs13481620*	12	1751.888	26	rs13481620*	12	1751.888	68
rs8273308*	12	1753.25	23	rs8273308*	12	1753.25	22
rs13482947	17	2246.618	16				

Table 5(a). Confirmed female biased genes. Genes that showed to be more than 2 fold higher expressed in females than in males in at least two previous studies, as well as in this current study. (GO: Gene Ontology)

Gene Symbol	Chromosome	Description	GO Biological Process
Abcd2	15	ATP-binding cassette, subfamily D2	Transport
Acot3	12	Acyl-CoA thioesterase 3	Acyl-CoA metabolism
Akr l b7	6	Aldo-keto reductase 1 B7	Cellular lipid metabolism
BCO 14805	19	cDNA sequence BC014805	Transport
BC089597	10	cDNA sequence BC089597	Metabolism
Ccnd 1	7	Cyclin D1	Regulation of cell cycle
Cyp2b 10	7	Cytochrome P450 2b10	Electron transport
Cyp2b 13	7	Cytochrome P450 2bl 3	Electron transport
Cyp2b9	7	Cytochrome P450 2b9	Electron transport
Cyp3a 16	5	Cytochrome	Electron transport
Cyp4a 10	4	Cytochrome	Electron transport
Cyp4 a 14	4	Cytochrome P450 4a14	Electron transport
Fmo3	1	Flavin-containing monooxygenase 3	Electron transport
Hao3	3	Hydroxyacid oxidase 3	Electron transport
Hexb	13	Hexosaminidase B	Calcium ion homeostasis
Mind2	5	Monocyte to macrophage differentiation-associated 2	Cytolysis
Npall	5	Nicotinamide N-methyhransferase	
Prlr	15	Prolactin receptor	Nucleotide catabolism
Prom 1	5	Prominin 1	Steroid biosynthesis

Rtn4	11	Reticulon 4	Regulation of cell migration
Serpinb 1 a	13	Serine (or cysteine) peptidase inhibitor	Regulation of protein catabolism
Slco 1 a4	6	Solute carrier family, member la4	Organic anion transport
Sult2a2	7	Sulfotransferase family 2A2	Steroid metabolism
Sult3a 1	10	Sulfotransferase family 3A1	
Tox	4	Thymocyte selection-associated HMG box gene	Regulation of transcription
Vldlr	19	Very-low-density lipoprotein receptor	Lipid metabolism

Table 5(b) Confirmed male biased genes. Genes that showed to be more than 2 fold higher expressed in males than in females in at least two previous studies, as well as in this current study.

| Symbol	Chromosome	Description	Biological Process
2810439FO2Rik	18	RIKEN cDNA 2810439F02 gene	
Abcg2	6	ATP-binding cassette, subfamily G2	Transport
Cml4	6	Canello-like 4	Negative regulation of cell adhesion
Cyp4a 12	4	Cytochrome P450, 4a12	Electron transport
Ddx3y	Y	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3	
Egfr	11	Epidermal growth factor receptor	Signal transduction
Eif2s3y	Y	Eukaryotic translation initiation factor 2, subunit 3	Protein biosynthesis
Elovl3	19	Elongation of very-long-chain fatty acids-like 3	Fatty-acid biosynthesis
Gstp1	19	Glutathione S-transferase, pi 1	Glutathione metabolism
Hsd3b5	3	Hydroxysteroid dehydrogenase-5	Steroid biosynthesis
Jaridld	Y	Jumonji, AT-rich interactive domain 1D	Regulation of transcription
Nudt7	8	Major urinary protein 4	Coenzyme A catabolism
Omd	13	Osteomodulin	Cell adhesion
Scara5	14	Scavenger receptor class A5	
Slco 1 a1	6	Solute carrier transpoter family 1a	Organic anion transport
Susd4	1	Sushi domain containing 4	
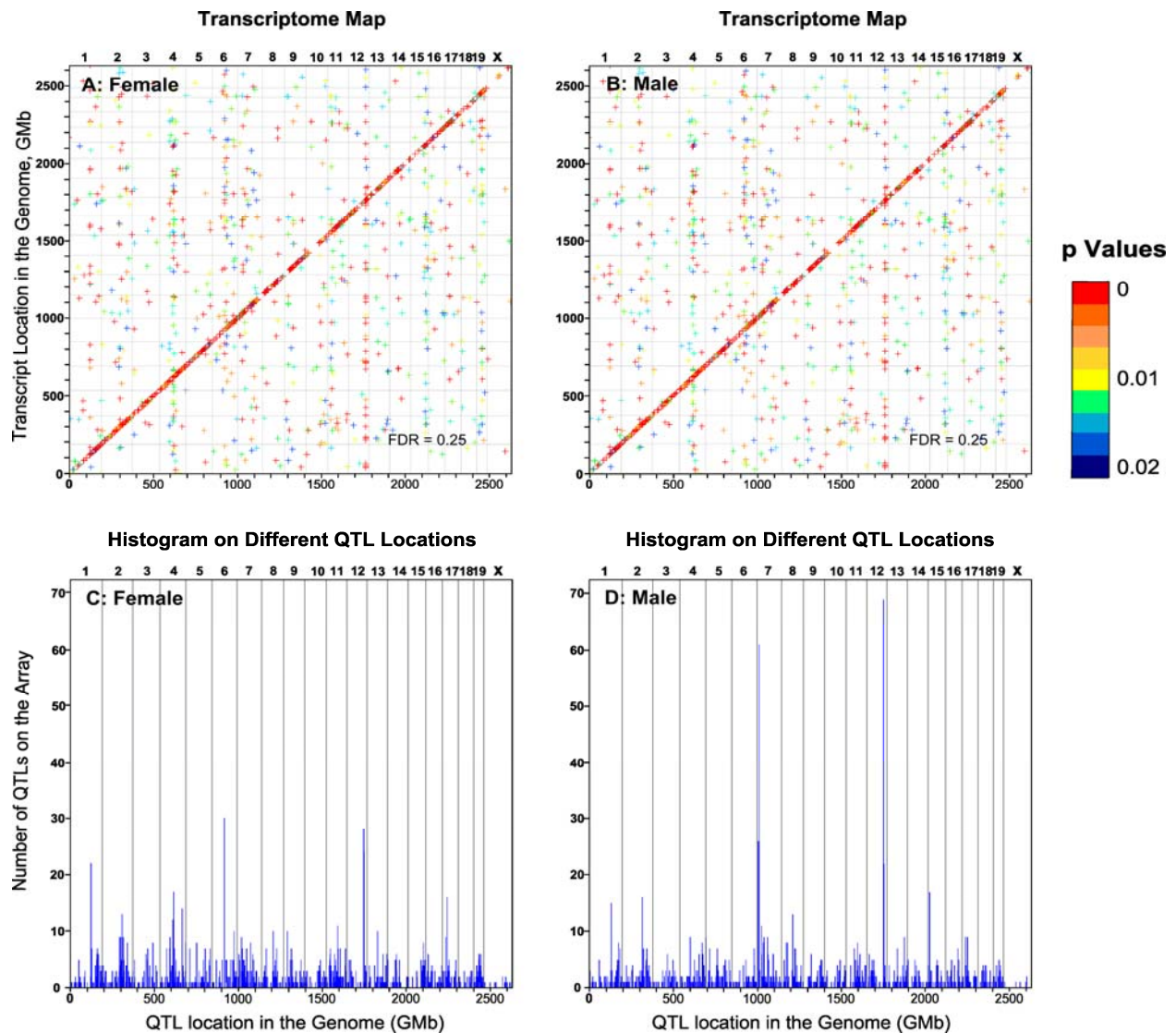
Figure 1. Sex-specific transcriptome maps reveal differences and similarities in the genetic regulation of gene expression in mouse liver. The female (A) and male (B) transcriptome maps are shown. Genomic location of each SNP marker (horizontal axis) and each transcript (vertical axis) are plotted. Each cross represents the location of the maximum QTL for a particular gene. Locally regulated (cis-eQTL) genes are located along the 45 degree lines while the vertical lines correspond to the loci which regulate distant (trans-eQTL) genes. The color of each symbol corresponds to the significance of the QTL (color bar). (C, D) Histogram counts of the number of transcripts regulated at each SNP marker.
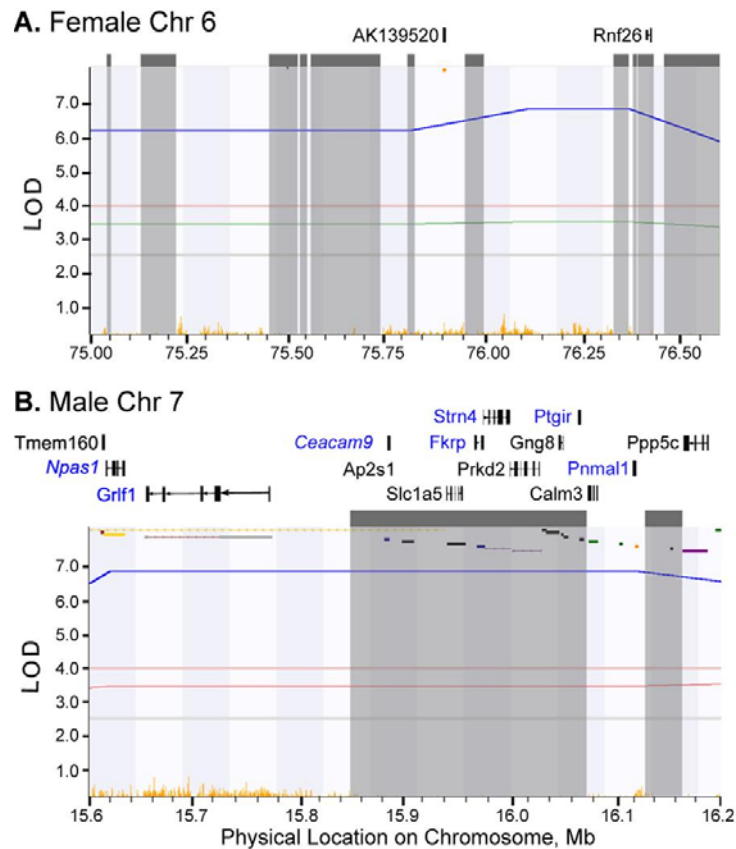
Figure 2. Master regulator loci in female (A) and male (B) mouse liver. (A) Chromosome 6 contains a female-specific QTL hotspot, with 28 transcripts regulated by this locus (FDR <= 0.25, average LOD score = 4.98). The positive additive coefficient (slim green line) indicates that DBA2J increases the trait value. Identical-by-descent regions between DBA/2J and C57BL/6J strains are shaded in gray. The top panel shows annotated genes located in the QTL region. Genes which have non-synonymous coding SNPs are labeled in blue, local-regulated genes in italics and trans-regulated genes in plain text. The gray location markers indicate that the gene was not represented on the array. The pink and the gray lines represent the significant and suggestive threshold, which are generated by permutation tests. (B) A region on chromosome 7 contains a male-specific QTL hotspot, with 55 transcripts regulated by this locus (FDR <= 0.25, average LOD score 5.05). The negative additive coefficient (the slim red line) indicates that C57B6J increases the trait value.
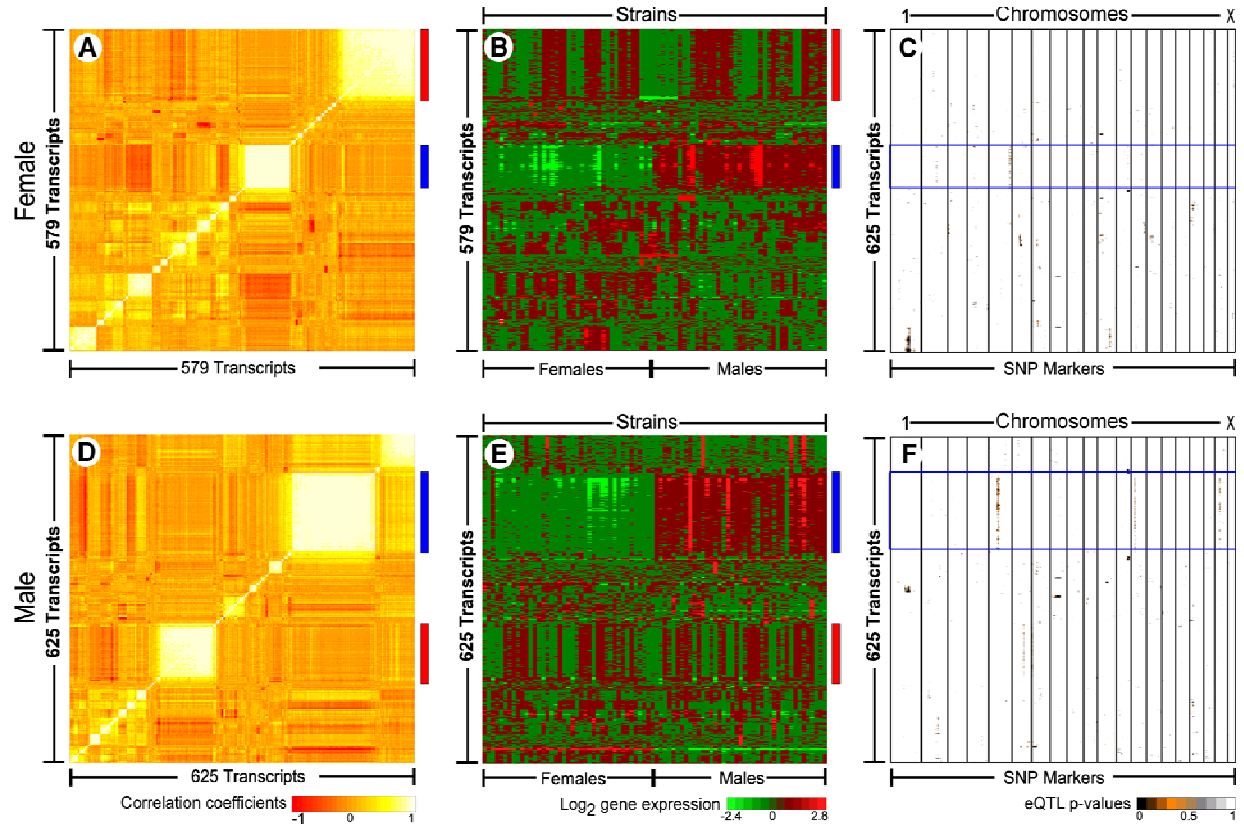
Figure 3. Co-correlation analysis of gene expression reveals sex-specific 'cliques.' Genes with high ($|r_{pearson}|>0.9$) correlation in female (A-C) and male (D-F) liver transcriptome were selected. (A, D) Clustering of transcripts based on Pearson correlation coefficients. (B, E) Two dimensional hierarchical clustering of the expression of the highly correlated genes in each strain/sex. The genes were shown in the same order along the vertical axis as in A and D. (C, F) Hierarchical clustering diagram of the genetic control of gene expression of the highly correlated genes in each strain/sex. Genetic loci associated with gene expression phenotypes with small p-values were mapped onto the mouse genome.
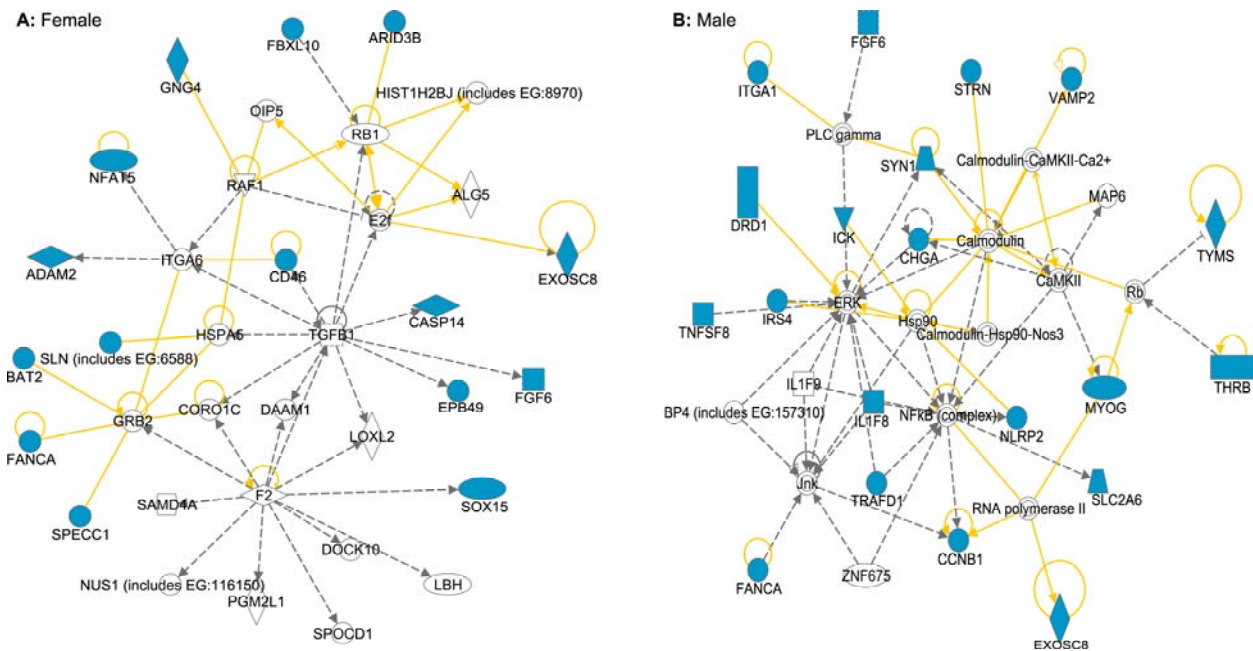
Figure 4: The top functional networks (Ingenuity®) identified for the highly correlated genes represented by the blue bars in figure 3. (A) The top network ("Cell-cell signaling and interaction") identified in female mice, with an ingenuity score of 36. (B) The top network ("Molecular transport") identified in male mice, with an ingenuity score of 42. The genes in blue are in the set of highly correlated genes and genes in plain color are curated by Ingenuity to form a network. Yellow lines indicate direct interactions. Ellipse, square, triangle, trapezoikd, lozenge and circle represent transcription regulator, cytokine, kinase, transporter, enzyme and other molecules, respectively.

References

1. **Waxman DJ**. Interactions of hepatic cytochromes P-450 with steroid hormones. Regioselectivity and stereospecificity of steroid metabolism and hormonal regulation of rat P-450 enzyme expression. *Biochem Pharmacol* 37: 71-84, 1988.

2. **Shapiro BH, Agrawal AK, Pampori NA**. Gender differences in drug metabolism regulated by growth hormone. *Int J Biochem Cell Biol* 27: 9-20, 1995.

3. **Greenblatt DJ, von Moltke LL**. Gender has a small but statistically significant effect on clearance of CYP3A substrate drugs. *J Clin Pharmacol* 48: 1350-1355, 2008.

4. **Muller C**. Liver, alcohol and gender. *Wien Med Wochenschr* 156: 523-526, 2006.

5. **Kim YC, Yim HK, Jung YS, Park JH, Kim SY**. Hepatic injury induces contrasting response in liver and kidney to chemicals that are metabolically activated: role of male sex hormone. *Toxicol Appl Pharmacol* 223: 56-65, 2007.

6. **Holt MP, Ju C**. Mechanisms of drug-induced liver injury. *AAPS J* 8: E48-E54, 2006.

7. **Miller MA**. Gender-based differences in the toxicity of pharmaceuticals--the Food and Drug Administration's perspective. *Int J Toxicol* 20: 149-152, 2001.

8. **Ostapowicz G, Fontana RJ, Schiodt FV, Larson A, Davern TJ, Han SH, McCashland TM, Shakil AO, Hay JE, Hynan L, Crippin JS, Blei AT, Samuel G, Reisch J, Lee WM**. Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. *Ann Intern Med* 137: 947-954, 2002.

9. **Shimizu I, Kohno N, Tamaki K, Shono M, Huang HW, He JH, Yao DF**. Female hepatology: favorable role of estrogen in chronic liver disease with hepatitis B virus infection. *World J Gastroenterol* 13: 4295-4305, 2007.

10. **Fattovich G, Stroffolini T, Zagni I, Donato F**. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology* 127: S35-S50, 2004.

11. **El Serag HB, Rudolph KL**. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 132: 2557-2576, 2007.

12. **Alonso LC, Rosenfield RL**. Oestrogens and puberty. *Best Pract Res Clin Endocrinol Metab* 16: 13-30, 2002.

13. **Khosla S, Melton LJ, III, Atkinson EJ, O'Fallon WM, Klee GG, Riggs BL**. Relationship of serum sex steroid levels and bone turnover markers with bone mineral

density in men and women: a key role for bioavailable estrogen. *J Clin Endocrinol Metab* 83: 2266-2274, 1998.

14. **Hahn S**. The role of TAFs in RNA polymerase II transcription. *Cell* 95: 579-582, 1998.

15. **Berk AJ**. Activation of RNA polymerase II transcription. *Curr Opin Cell Biol* 11: 330-335, 1999.

16. **Boonyaratanakornkit V, Edwards DP**. Receptor mechanisms mediating non-genomic actions of sex steroids. *Semin Reprod Med* 25: 139-153, 2007.

17. **Pettersson K, Gustafsson JA**. Role of estrogen receptor beta in estrogen action. *Annu Rev Physiol* 63: 165-192, 2001.

18. **Nagasue N, Ito A, Yukaya H, Ogawa Y**. Androgen receptors in hepatocellular carcinoma and surrounding parenchyma. *Gastroenterology* 89: 643-647, 1985.

19. **Chowen JA, Frago LM, Argente J**. The regulation of GH secretion by sex steroids. *Eur J Endocrinol* 151 Suppl 3: U95-100, 2004.

20. **Waxman DJ, Pampori NA, Ram PA, Agrawal AK, Shapiro BH**. Interpulse interval in circulating growth hormone patterns regulates sexually dimorphic expression of hepatic cytochrome P450. *Proc Natl Acad Sci U S A* 88: 6868-6872, 1991.

21. **Lahn BT, Page DC**. Functional coherence of the human Y chromosome. *Science* 278: 675-680, 1997.

22. **Dobyns WB, Filauro A, Tomson BN, Chan AS, Ho AW, Ting NT, Oosterwijk JC, Ober C**. Inheritance of most X-linked traits is not dominant or recessive, just X-linked. *Am J Med Genet A* 129A: 136-143, 2004.

23. **Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL**. Sex-dependent gene expression and evolution of the Drosophila transcriptome. *Science* 300: 1742-1745, 2003.

24. **Reinke V, Gil IS, Ward S, Kazmer K**. Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. *Development* 131: 311-323, 2004.

25. **Santos EM, Kille P, Workman VL, Paull GC, Tyler CR**. Sexually dimorphic gene expression in the brains of mature zebrafish. *Comp Biochem Physiol A Mol Integr Physiol* 149: 314-324, 2008.

26. **Nishida Y, Yoshioka M, St Amand J**. Sexually dimorphic gene expression in the hypothalamus, pituitary gland, and cortex. *Genomics* 85: 679-687, 2005.

27. **Reinius B, Saetre P, Leonard JA, Blekhman R, Merino-Martinez R, Gilad Y, Jazin E**. An evolutionarily conserved sexual signature in the primate brain. *PLoS Genet* 4: e1000100-2008.

28. **Pampori NA, Shapiro BH**. Feminization of hepatic cytochrome P450s by nominal levels of growth hormone in the feminine plasma profile. *Mol Pharmacol* 50: 1148-1156, 1996.

29. **Agrawal AK, Shapiro BH**. Differential expression of gender-dependent hepatic isoforms of cytochrome P-450 by pulse signals in the circulating masculine episodic growth hormone profile of the rat. *J Pharmacol Exp Ther* 292: 228-237, 2000.

30. **Lichanska AM, Waters MJ**. How growth hormone controls growth, obesity and sexual dimorphism. *Trends Genet* 24: 41-47, 2008.

31. **Ahluwalia A, Clodfelter KH, Waxman DJ**. Sexual dimorphism of rat liver gene expression: regulatory role of growth hormone revealed by deoxyribonucleic Acid microarray analysis. *Mol Endocrinol* 18: 747-760, 2004.

32. **Yang X, Schadt EE, Wang S, Wang H, Arnold AP, Ingram-Drake L, Drake TA, Lusis AJ**. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res* 16: 995-1004, 2006.

33. **Wade CM, Daly MJ**. Genetic variation in laboratory mice. *Nat Genet* 37: 1175-1180, 2005.

34. **Treadwell JA**. Integrative strategies to identify candidate genes in rodent models of human alcoholism. *Genome* 49: 1-7, 2006.

35. **Gill K, Liu Y, Deitrich RA**. Voluntary alcohol consumption in BXD recombinant inbred mice: relationship to alcohol metabolism. *Alcohol Clin Exp Res* 20: 185-190, 1996.

36. **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* 1: 1-14, 1986.

37. **Lee GH, Bennett LM, Carabeo RA, Drinkwater NR**. Identification of hepatocarcinogen-resistance genes in DBA/2 mice. *Genetics* 139: 387-395, 1995.

38. **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* 54: 1191-1199, 2005.

39. **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* 28: 599-605, 1993.

40. **Brem RB, Yvert G, Clinton R, Kruglyak L**. Genetic dissection of transcriptional regulation in budding yeast. *Science* 296: 752-755, 2002.

41. **Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH**. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422: 297-302, 2003.

42. **Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, Hsu HC, Mountz JD, Baldwin NE, Langston MA, Threadgill DW, Manly KF, Williams RW**. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet* 37: 233-242, 2005.

43. **Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T, Su AI, Vellenga E, Wang J, Manly KF, Lu L, Chesler EJ, Alberts R, Jansen RC, Williams RW, Cooke MP, de Haan G**. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat Genet* 37: 225-232, 2005.

44. **Wang X, Paigen B**. Genetics of variation in HDL cholesterol in humans and mice. *Circ Res* 96: 27-42, 2005.

45. **Grisel JE, Metten P, Wenger CD, Merrill CM, Crabbe JC**. Mapping of quantitative trait loci underlying ethanol metabolism in BXD recombinant inbred mouse strains. *Alcohol Clin Exp Res* 26: 610-616, 2002.

46. **Clodfelter KH, Miles GD, Wauthier V, Holloway MG, Zhang X, Hodor P, Ray WJ, Waxman DJ**. Role of STAT5a in regulation of sex-specific gene expression in female but not male mouse liver revealed by microarray analysis. *Physiol Genomics* 31: 63-74, 2007.

47. **Amador-Noguez D, Zimmerman J, Venable S, Darlington G**. Gender-specific alterations in gene expression and loss of liver sexual dimorphism in the long-lived Ames dwarf mice. *Biochem Biophys Res Commun* 332: 1086-1100, 2005.

48. **Rinn JL, Rozowsky JS, Laurenzi IJ, Petersen PH, Zou K, Zhong W, Gerstein M, Snyder M**. Major molecular differences between mammalian sexes are involved in drug metabolism and renal function. *Dev Cell* 6: 791-800, 2004.

49. **Su WL, Modrek B, Guhathakurta D, Edwards S, Shah JK, Kulkarni AV, Russell A, Schadt EE, Johnson JM, Castle JC**. Exon and junction microarrays detect widespread mouse strain- and sex-bias expression differences. *BMC Genomics* 9: 273-2008.

50. **Sornson MW, Wu W, Dasen JS, Flynn SE, Norman DJ, O'Connell SM, Gukovsky I, Carriere C, Ryan AK, Miller AP, Zuo L, Gleiberman AS, Andersen B, Beamer WG, Rosenfeld MG**. Pituitary lineage determination by the Prophet of Pit-1 homeodomain factor defective in Ames dwarfism. *Nature* 384: 327-333, 1996.

51. **Chandrashekar V, Bartke A**. Induction of endogenous insulin-like growth factor-I secretion alters the hypothalamic-pituitary-testicular function in growth hormone-deficient adult dwarf mice. *Biol Reprod* 48: 544-551, 1993.

52. **Clodfelter KH, Holloway MG, Hodor P, Park SH, Ray WJ, Waxman DJ**. Sex-dependent liver gene expression is extensive and largely dependent upon signal transducer and activator of transcription 5b (STAT5b): STAT5b-dependent activation of male genes and repression of female genes revealed by microarray analysis. *Mol Endocrinol* 20: 1333-1351, 2006.

53. **Bauer-Hofmann R, Klimek F, Buchmann A, Muller O, Bannasch P, Schwarz M**. Role of mutations at codon 61 of the c-Ha-*ras* gene during diethylnitrosamine-induced hepatocarcinogenesis in C3H/He mice. *Mol Carcinogen* 6: 60-67, 1992.

54. **Lin JX, Leonard WJ**. The role of Stat5a and Stat5b in signaling by IL-2 family cytokines. *Oncogene* 19: 2566-2576, 2000.

55. **Grimley PM, Dong F, Rui H**. Stat5a and Stat5b: fraternal twins of signal transduction and transcriptional activation. *Cytokine Growth Factor Rev* 10: 131-157, 1999.

56. **Frank SJ**. Growth hormone signalling and its regulation: preventing too much of a good thing. *Growth Horm IGF Res* 11: 201-212, 2001.

57. **Enomoto N, Takei Y, Yamashina S, Ikejima K, Suzuki S, Kitamura T, Sato N**. [Gender difference in alcoholic liver injury]. *Nihon Arukoru Yakubutsu Igakkai Zasshi* 39: 163-167, 2004.

58. **Wang S, Yehya N, Schadt EE, Wang H, Drake TA, Lusis AJ**. Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genet* 2: e15-2006.

59. **Tullis KM, Krebs CJ, Leung JY, Robins DM**. The regulator of sex-limitation gene, rsl, enforces male-specific liver gene expression by negative regulation. *Endocrinology* 144: 1854-1860, 2003.

60. **Wiwi CA, Waxman DJ**. Role of hepatocyte nuclear factors in growth hormone-regulated, sexually dimorphic expression of liver cytochromes P450. *Growth Factors* 22: 79-88, 2004.

61. **Su Z, Li Y, James JC, McDuffie M, Matsumoto AH, Helm GA, Weber JL, Lusis AJ, Shi W**. Quantitative Trait Locus Analysis of Atherosclerosis in an Intercross Between C57BL/6 and C3H Mice Carrying the Mutant Apolipoprotein E Gene. *Genetics* 172: 1799-1807, 2006.

62. **Isensee J, Ruiz NP**. Sexually dimorphic gene expression in mammalian somatic tissue. *Gend Med* 4 Suppl B: S75-S95, 2007.

63. **Farrall M**. Quantitative genetic variation: a post-modern view. *Hum Mol Genet* 13 Spec No 1: R1-R7, 2004.

64. **Shi C, Uzarowska A, Ouzunova M, Landbeck M, Wenzel G, Lubberstedt T**. Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint x Flint maize recombinant inbred line population. *BMC Genomics* 8: 22-2007.

65. **Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S, Phillips JW, Sachs A, Schadt EE**. Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet* 75: 1094-1105, 2004.

66. **Wu C, Delano DL, Mitro N, Su SV, Janes J, McClurg P, Batalov S, Welch GL, Zhang J, Orth AP, Walker JR, Glynne RJ, Cooke MP, Takahashi JS, Shimomura K, Kohsaka A, Bass J, Saez E, Wiltshire T, Su AI**. Gene set enrichment in eQTL data identifies novel annotations and pathway regulators. *PLoS Genet* 4: e1000070-2008.

67. **Gatti D, Maki A, Chesler EJ, Kirova R, Kosyk O, Lu L, Manly KF, Williams RW, Perkins A, Langston MA, Threadgill DW, Rusyn I**. Genome-level analysis of genetic regulation of liver gene expression networks. *Hepatology* 46: 548-557, 2007.

68. **Barry WT, Nobel AB, Wright FA**. Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics* 21: 1943-1949, 2005.

69. **Gatti DM, Sypa M, Rusyn I, Wright FA, Barry WT**. SAFEGUI: resampling-based tests of categorical significance in gene expression data made easy. *Bioinformatics* 25: 541-542, 2009.

70. **Voy BH, Scharff JA, Perkins AD, Saxton AM, Borate B, Chesler EJ, Branstetter LK, Langston MA**. Extracting gene networks for low-dose radiation using graph theoretical algorithms. *PLoS Comput Biol* 2: e89-2006.

71. **Manly K**. QTLReaper. 2003.

72. **Churchill GA, Doerge RW**. Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963-971, 1994.

73. **Wang J, Williams RW, Manly KF**. WebQTL: web-based complex trait analysis. *Neuroinformatics* 1: 299-308, 2003.

74. **Szatkiewicz JP, Beane GL, Ding Y, Hutchins L, Pardo-Manuel d, V, Churchill GA**. An imputed genotype resource for the laboratory mouse. *Mamm Genome* 19: 199-208, 2008.

75. **R Development Core Team**. R: A Language and Environment for Statistical Computing. 2006.

76. **Zhang Q, McMillan L, Pardo-Manuel de Villena F, Threadgill DW, Wang W**. Inferring genome-wide mosaic structure. *Proceedings of the 14th Pacific Symposium on Biocomputing (PSB)* 14: 150-161, 2009.

77. **Ho Sui SJ, Mortimer JR, Arenillas DJ, Brumm J, Walsh CJ, Kennedy BP, Wasserman WW**. oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res* 33: 3154-3164, 2005.

78. **Ho Sui SJ, Fulton DL, Arenillas DJ, Kwon AT, Wasserman WW**. oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Res* 35: W245-W252, 2007.

79. **Doss S, Schadt EE, Drake TA, Lusis AJ**. Cis-acting expression quantitative trait loci in mice. *Genome Res* 15: 681-691, 2005.

80. **Rockman MV, Kruglyak L**. Genetics of global gene expression. *Nat Rev Genet* 7: 862-872, 2006.

81. **Alberts R, Terpstra P, Bystrykh LV, de Haan G, Jansen RC**. A statistical multiprobe model for analyzing cis and trans genes in genetical genomics experiments with short-oligonucleotide arrays. *Genetics* 171: 1437-1439, 2005.

82. **Gatti D, Maki A, Chesler EJ, Kirova R, Kosyk O, Lu L, Manly KF, Williams RW, Perkins A, Langston MA, Threadgill DW, Rusyn I**. Genome-level analysis of genetic regulation of liver gene expression networks. *Hepatology* 46: 548-557, 2007.

83. **Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MF, Fisher EM**. Genealogies of mouse inbred strains. *Nat Genet* 24: 23-25, 2000.

84. **Svenson KL, Von Smith R, Magnani PA, Suetin HR, Paigen B, Naggert JK, Li R, Churchill GA, Peters LL**. Multiple Trait Measurements in 43 Inbred Mouse Strains Captures the Phenotypic Diversity Characteristic of Human Populations. *J Appl Physiol* 2007.

85. **Wiwi CA, Gupte M, Waxman DJ**. Sexually dimorphic P450 gene expression in liver-specific hepatocyte nuclear factor 4alpha-deficient mice. *Mol Endocrinol* 18: 1975-1987, 2004.

86. **Mulder TP, Verspaget HW, Sier CF, Roelofs HM, Ganesh S, Griffioen G, Peters WH**. Glutathione S-transferase pi in colorectal tumors is predictive for overall survival. *Cancer Res* 55: 2696-2702, 1995.

87. **Hatayama I, Satoh K, Sato K**. Developmental and hormonal regulation of the major form of hepatic glutathione S-transferase in male mice. *Biochem Biophys Res Commun* 140: 581-588, 1986.

88. **Koukouritaki SB, Simpson P, Yeung CK, Rettie AE, Hines RN**. Human hepatic flavin-containing monooxygenases 1 (FMO1) and 3 (FMO3) developmental expression. *Pediatr Res* 51: 236-243, 2002.

89. **Cherrington NJ, Cao Y, Cherrington JW, Rose RL, Hodgson E**. Physiological factors affecting protein expression of flavin-containing monooxygenases 1, 3 and 5. *Xenobiotica* 28: 673-682, 1998.

90. **Cashman JR**. Human flavin-containing monooxygenase: substrate specificity and role in drug metabolism. *Curr Drug Metab* 1: 181-191, 2000.

91. **Teh CH, Lam KK, Loh CC, Loo JM, Yan T, Lim TM**. Neuronal PAS domain protein 1 is a transcriptional repressor and requires arylhydrocarbon nuclear translocator for its nuclear localization. *J Biol Chem* 281: 34617-34629, 2006.

92. **Ohsawa S, Hamada S, Kakinuma Y, Yagi T, Miura M**. Novel function of neuronal PAS domain protein 1 in erythropoietin expression in neuronal cells. *J Neurosci Res* 79: 451-458, 2005.

93. **Erbel-Sieler C, Dudley C, Zhou Y, Wu X, Estill SJ, Han T, Diaz-Arrastia R, Brunskill EW, Potter SS, McKnight SL**. Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proc Natl Acad Sci U S A* 101: 13648-13653, 2004.

94. **Teh CH, Loh CC, Lam KK, Loo JM, Yan T, Lim TM**. Neuronal PAS domain protein 1 regulates tyrosine hydroxylase level in dopaminergic neurons. *J Neurosci Res* 85: 1762-1773, 2007.

95. **Zhou YD, Barnard M, Tian H, Li X, Ring HZ, Francke U, Shelton J, Richardson J, Russell DW, McKnight SL**. Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. *Proc Natl Acad Sci U S A* 94: 713-718, 1997.

96. **Finkenzeller D, Fischer B, McLaughlin J, Schrewe H, Ledermann B, Zimmermann W**. Trophoblast cell-specific carcinoembryonic antigen cell adhesion molecule 9 is not required for placental development or a positive outcome of allotypic pregnancies. *Mol Cell Biol* 20: 7140-7145, 2000.

97. **Jiang W, Betson M, Mulloy R, Foster R, Levay M, Ligeti E, Settleman J**. p190A RhoGAP is a glycogen synthase kinase-3-beta substrate required for polarized cell migration. *J Biol Chem* 283: 20978-20988, 2008.

98. **Tikoo A, Czekay S, Viars C, White S, Heath JK, Arden K, Maruta H**. p190-A, a human tumor suppressor gene, maps to the chromosomal region 19q13.3 that is reportedly deleted in some gliomas. *Gene* 257: 23-31, 2000.

99. **Wolfe CJ, Kohane IS, Butte AJ**. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics* 6: 227-2005.

100. **Zysow BR, Kauser K, Lawn RM, Rubanyi GM**. Effects of estrus cycle, ovariectomy, and treatment with estrogen, tamoxifen, and progesterone on apolipoprotein(a) gene expression in transgenic mice. *Arterioscler Thromb Vasc Biol* 17: 1741-1745, 1997.

101. **Silberstein GB, Van Horn K, Hrabeta-Robinson E, Compton J**. Estrogen-triggered delays in mammary gland gene expression during the estrous cycle: evidence for a novel timing system. *J Endocrinol* 190: 225-239, 2006.