Boston University Theses & Dissertations

Boston University Theses & Dissertations

2018

# Role of long noncoding RNAs and genetic variants in the regulation of sex-specific gene expression patterns in mouse liver

https://hdl.handle.net/2144/33194 Downloaded from DSpace Repository, DSpace Institution's institutional repository

# BOSTON UNIVERSITY

# GRADUATE SCHOOL OF ARTS AND SCIENCES

# AND

# COLLEGE OF ENGINEERING

Dissertation

# ROLE OF LONG NONCODING RNAS AND GENETIC VARIANTS IN THE REGULATION OF SEX-SPECIFIC GENE EXPRESSION PATTERNS IN MOUSE LIVER

by

# TISHA MELIA

B.S., Michigan State University, 2005 M.Sc., Technische Universiteit Delft, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2018

© 2018 by Tisha Melia All rights reserved Approved by

First Reader

David J. Waxman, Ph.D. Professor of Biology, Professor of Medicine

Second Reader

Trevor Siggers, Ph.D. Assistant Professor of Biology Dedicated to the memory of my father,

who always believed in me

#### Acknowledgments

I thank God for giving me the strength and determination to complete this PhD. I sincerely hope that any knowledge that I gained during my PhD could be useful to other people, especially to the people of Indonesia.

I thank my advisor, David, for the opportunity and guidance during the last 5.5 years. I appreciate all of your efforts to mentor me.

I thank Stefano for chairing my committee, and making sure that I could graduate as soon as possible. I thank Dr. Trevor Siggers for agreeing to be my second reader. I also thank the rest of my committee: Dr. Thomas Tullius and Dr. Mark Kon for giving me scientific feedbacks to improve my research.

I thank past and current members of the Waxman Lab for years of continued support and sympathy: Gracia Bonilla, Dr. Christina Hao, Christine Nykyforchyn, Kritika Karri, George Steinhardt, Andy Rampersaud, Cindy Marmol, Dana Lau, Aram Shin, Bryan Matthews, Dr. Nick Lodato, Dr. Marie Jordan, Dr. Junjie Wu, Dr. Li Jia, Dr. Mona Connerney, Dr. Sasha Suvorov, Dr. Bin Du and Chong Sheng. I particularly would like to thank Gracia, whom I always shared cubicle with since my first year at BU; There is a reason why people called our workspace "the fun cubicle". Christina, thank you for always understanding me when no words can describe my feelings. George and Andy, thank you for always lifting up my mood. Christine, thank you for always stopping by tochat. Kritika, thank you for all those food and encouragements.I thank all of my friends in Boston for being my surrogate family: Amrita, Jaeyoon,Aissatou, Trisia, Ayeyo, Maryam, Faduma, Fatma, Ameenah, Samiya and her two sisters,Khadra, Mama and Baba, as well as Mbak Dewi and family.

I thank the Bioinformatics staffs for helping me with all the paperwork: Caroline, Johanna, Dave and Mary Ellen.

To my three brothers, thank you for the most fun childhood a sibling can ask.

To my husband, Uda, without your endless support, I could not have done this. Thank you for all of your sacrifices. Thank you for sharing our two beautiful children with me.

To my children: Maryam and Yahya, you are both the reason that I got up in the morning every day to finish this dissertation.

To my mom, thank you is not enough to describe how much I am grateful for your love and support. You are the strongest woman I know. I love you.

To my beloved father, thank you for always believing in me. I know you would be happy to know that I finished this PhD. Thank you for being the best dad. Rest in peace.

# ROLE OF LONG NONCODING RNAS AND GENETIC VARIANTS IN THE REGULATION OF SEX-SPECIFIC GENE EXPRESSION PATTERNS IN MOUSE LIVER

# TISHA MELIA

Boston University Graduate School of Arts and Sciences

and

College of Engineering, 2018

Major Professor: David Waxman, Professor of Biology and Professor of Medicine

# Abstract

Sex biased expression characterizes ~1,000 genes in mammalian liver, and impart sex differences in metabolism and disease susceptibility. The sex-dependent temporal patterns of pituitary growth hormone (GH) secretion, pulsatile in males and more continuous in females, are known to sex-differentially activate transcriptional regulators (TFs), leading to widespread sex-differences in the mouse liver transcriptome. This thesis elucidates sex-biased gene expression patterns in the following studies. Gene structures, expression patterns and species conservation are characterized for ~15,000 liverexpressed intergenic long noncoding RNAs (lncRNAs), many of which are novel. Analysis of intergenic lncRNA promoters revealed unexpected high conservation and significant enrichment of TF binding compared to protein-coding promoters. A subset of intergenic lncRNAs showed strong sex-specific and GH-dependent gene expression, and whose transcription was tightly correlated with the surrounding chromatin environment and TF binding patterns. The pervasive role of genetic factors to regulate sex-biased genes was revealed by analyzing livers with matched genotype and gene expression data from Diversity Outbred (DO) mice, an outbred population with high natural allelic variance derived from eight inbred strains. Significant associations between genetic variants and gene expression (eQTLs) were identified, including many eQTLs with a strong sex-biased association. Remarkably, a large fraction of these sex-biased eQTLs were linked to either gain or loss of sex-specific gene expression in the DO founder strain predicted to be regulated by the eQTL. Thus, genetic factors are a major contributor to the variability of sex-biased genes, which has important consequences related to the individual variability of liver phenotypes with known sex-differences. Natural genetic perturbations in DO mice were leveraged to identify candidate lncRNAs that may regulate hypophysectomy (hypox) responsiveness. Co-regulated protein-coding gene clusters were discovered based on gene expression correlations across DO mouse livers, many of which are enriched for distinct hypox response classes. LncRNAs whose expression showed unexpected significant negative correlation with protein-coding gene clusters enriched for genes of the opposite-sex bias and inverse hypox class were hypothesized to play negative regulatory role. In sum, these studies expand the characterization of the sex-biased hepatic transcriptome and reveal contributions of genetic factors to the regulation of sex bias in mammalian liver.

# **Table of Contents**

Acknowledgmentsv
Abstractvii
Table of Contents ix
List of Tables xi
List of Figuresxii
List of Abbreviations xv
Chapter 1 Introduction
1.A Summary
1.B GH secretion patterns regulate liver sex-biased genes
1.C GH-dependent TFs
1.D Classification of sex-biased genes in response to hypophysectomy
1.E GH-dependent sex differences in liver epigenome
1.F Sex-biased genes during mouse liver development7
Chapter 2 Hepatic lincRNAs: high promoter conservation and dynamic, sex-dependent
transcriptional regulation by growth hormone
2.A Abstract
2.B Introduction
2.C Materials and methods
2.D Results

2.E Discussion	46
2.F Acknowledgements	53
Chapter 3 Genetic factors increase the variability of sex-specific gene patterns in	
Diversity Outbred mouse liver	68
3.A Abstract	68
3.B Introduction	69
3.C Materials and methods	72
3.D Results	81
3.E Discussion	94
Chapter 4 Sex-biased gene co-expression networks in Diversity Outbred mouse liver. 1	57
4.A Abstract	57
4.B Introduction	58
4.C Materials and methods	61
4.D Results	70
4.E Discussion	80
4.F Acknowledgements1	85
LIST OF JOURNAL ABBREVIATIONS	11
BIBLIOGRAPHY2	14
CURRICULUM VITAE	38

# List of Tables

Table 2-1. Exon and introns of novel lincRNA, RefSeq protein-coding, and RefSeq
noncoding genes
Table 2-2. Sex-specific mouse liver lincRNAs that respond to hypophysectomy (hypox)
Table 3-1. Number of autosomal eQTLs found when using all, male only or female only
samples
Table S3-1 to Table S3-8
Table 4-1. Enrichments of gene modules discovered in male, and separately, in female
liver for sex-specific and hypox-responsive genes
Table S4-2. List of sex-specific lncRNAs that are negatively correlated (FDR $< 0.001$ )
with protein-coding genes enriched ( $P < 0.01$ ) for inverse hypox classes
Table S4-1 to Table S4-7

# **List of Figures**

Figure 2-1. Liver-expressed lincRNAs. 56
Figure 2-2. Conservation of lincRNA promoters
Figure 2-3. Liver lincRNA species conservation and tissue specificity
Figure 2-4. GH regulation of sex-specific liver lincRNAs
Figure 2-5. Regulation of select sex-specific liver lincRNAs
Figure 2-6. Sex-specific TF binding and chromatin states at sex-specific lincRNAs 66
Figure 3-1. Protein-coding gene sex-specificity across DO founder strains 105
Figure 3-2. Autosomal eQTLs for liver-expressed genes
Figure 3-3. Autosomal eQTLs associated with sex-specific genes
Figure 3-4. Enrichment of DHS and TF binding sites that contain regulating strain-
specific SNPs/indels at different types of eQTL regions
Figure 3-5. Co-regulated gene clusters based on overlapping eQTL regions discovered in
all liver samples 113
Figure 3-6. Co-regulated gene clusters that show different gene members in male vs.
female liver 115
Figure S3-1. Schematic overview of how the SNP array and RNA-Seq data were
analyzed for each DO mouse liver samples117
Figure S3-2. High variability of the sex-specificity of gene expression in individual livers
of CD-1 mice and DO mouse founder strains for select male-specific and female-
specific genes
Figure S3-3. Sex-specific protein-coding genes across DO founder strains

Figure S3-4. Properties of autosomal eQTLs in DO mouse liver
Figure S3-5. Examples of genes with stronger genetic association in male or female liver.
Figure S3-6. Examples of sex-specific genes in categories 1-8, described in Fig. 3-3 128
Figure S3-7. Regression coefficients (top figure in each panel) and LOD scores (bottom
figure in each panel) for genes listed in Fig. 3-5A
Figure S3-8. Genome-wide LOD score plots for genes that are located in different
chromosomes in the co-regulated cluster depicted in Fig. 3-5A 150
Figure S3-9. Genome-wide LOD score plots the Sult2a1, Sult2a3 and Sult2a5 genes. 153
Figure S3-10. Strain-specific SNPs/indels are hypothesized to contribute to the loss of
sex-specific gene expression for <i>Cyp2b9</i> 154
Figure 4-1. Liver-expressed lncRNAs
Figure 4-2. Sex-specific lncRNAs showing significant regulation during postnatal liver
developmental and during a 24-hour light-dark cycle periods in male mouse liver.
Figure 4-3. Sex-specific gene clusters (gene modules) discovered in DO male mouse
liver
Figure 4-4. Sex-specific gene clusters discovered in DO female mouse liver 195
Figure 4-5. Overlap between clusters discovered in male vs. female liver for clusters that
are significantly enriched for sex-specific genes and/or hypox-response gene classes.

Figure 4-6. Schematic diagram of the eight proposed scenarios for negative regulatory		
role of sex-specific lncRNAs		
Figure 4-7. Distribution of hypox-response gene classes for intergenic, sex-specific		
lncRNAs whose expression in DO mouse liver is eitherpositively or negatively		
correlated with protein-coding gene modules that are enriched for one the four		
hypox-response classes		
Figure S4-1. Four hypox-response classes		
Figure S4-2. Pipeline to discover liver-expressed lncRNAs		
Figure S4-3. A Heatmap of 684 sex-specific lncRNAs in different RNA-Seq datasets. 205		
Figure S4-4. Overlap between clusters discovered in male vs. female liver		
Figure S4-5. Examples of gene modules that are negatively correlated with sex-specific		
lncRNAs (Table 4-2)		

# List of Abbreviations

BCL6	B-cell lymphoma 6
bp	Base pair
CUX2	Cut-like homeobox 2
DAVID	The Database for Annotation, Visualization and Integrated Discovery
DHS	DNase hypersensitive site(s)
DO	Diversity Outbred
eRNA	Enhancer RNA
ES	Enrichment score
eQTL	Expression quantitative trait loci
FPKM	
FPR	
GH	
H3K4me1	
H3K4me3	
H3K27ac	
H3K27me3	
H3K36me3	
HMM	
HNF	
Нурох	
KEGG	Kyoto Encyclopedia of Genes and Genomes

KRAB-ZPF	Krüppel-associated box zinc finger
lincRNA	Long intergenic non-coding RNA
IncRNA	Long non-coding RNA
LOD	Logarithm of the odds
MUGA	Mouse Universal Genotyping Array
ncRNA	Non-coding RNA
PCA	Principal component analysis
qPCR	Quantitative polymerase chain reaction
Rsl	
SNP	Single nucleotide polymorphism
STAT5	
TAD	
TES	Transcript end site
TF	Transcription factor
TSS	

#### **Chapter 1 Introduction**

### **1.A Summary**

The liver plays key physiological functions, including the metabolism of drug and lipid, as well as bile acid synthesis. Sex-differences are widespread in mammalian liver transcriptome, where, in mouse, there are >1,000 genes with significant sex-bias expression. Many liver sex-specific genes are implicated in the sex differences of clinically important variables, such as liver disease susceptibility and function. Growth hormone (GH) is known to be a key regulator of sex-biased genes through its secretion pattern that are different between the two sexes. In male, GH is secreted in pulsatile manner, whereas in female, it is secreted near continuously. Sex differences in GH plasma level sex-differentially activates STAT5b and other GH-dependent transcription factors (TF), propagating sex differences to the entire liver transcriptome. GH, by mechanisms that are not yet understood, impart sex-differences in the liver chromatin structure, histone marks and methylation status.

The goal of this thesis is to characterize the sex-biased liver transcriptome and its associated gene regulation mechanisms in mouse liver, which were accomplished by discovering liver-expressed long noncoding RNAs (lncRNAs; Chapter 2) and their putative regulatory roles (Chapter 4), as well as elucidating genetic factors altering gene expression of liver sex-biased genes (Chapter 3). Briefly, in Chapter 2, I discovered liver-expressed intergenic lncRNAs (lincRNAs), and characterized their gene expression patterns and conservation level. TF bindings, chromatin marks, and chromatin openings

data were integrated to show their tight correlation with the transcription of sex-biased lncRNAs. Chapter 3 delineates the role of genetic component in the regulation of sex-bias gene expression in the Diversity Outbred (DO) mouse, an outbred mouse population with high genetic diversity. In Chapter 4, I discovered gene co-expression modules in male, and separately in female liver, and showed their enrichment for biologically relevant traits. I used the identified gene modules to reveal broad characteristics of sex-bias gene regulation mechanisms, as well as a tool to generate hypothesis for lncRNA regulatory roles.

### **1.B GH secretion patterns regulate liver sex-biased genes**

Sex differences in gene expression have been observed in liver, kidney, brain, and other tissues in mammals (Gur et al. 1999, Rinn et al. 2004, Rinn and Snyder 2005). Sex specific genes in liver are perhaps the best studied, where many are implicated in sex differences in the metabolism of drugs, steroids and xenometabolites (Anthony and Berg 2002, Guengerich 2006, Klaassen and Aleksunes 2010, Liu et al. 2013, Zanger and Schwab 2013), as well as disease susceptibility, including hepatocellular carcinoma (Ruggieri et al. 2010, Baik et al. 2011), polygenic dyslipidemia (Bittner 2006) and coronary artery disease (Zhang et al. 2011). Rodent studies revealed growth hormone (GH) to be a major regulator of liver sex-biased genes (Jansson et al. 1985, Jaffe et al. 2002, Waxman and O'Connor 2006) through its sex-differential secretion patterns (Jaffe et al. 2002). In male rats, growth hormone is secreted in a highly pulsatile manner with peaks occurring every ~3.5 h; and growth hormone level between peaks drops to

undetectable amount. In contrast, the female growth hormone pattern in rats is continuous due to more frequent pulses (Tannenbaum and Martin 1976, Eden 1979). Similar sexdependent GH secretion patterns, albeit less striking, were observed in mouse (MacLeod et al. 1991), and to a much weaker degree in human (Pincus et al. 1996). The male GH secretion pattern, in rat, was imprinted by early exposure to neonatal androgens (Jansson et al. 1985). The sexual dimorphic growth hormone secretion patterns result in sex differences in the plasma GH level, which in turn activates transcriptional regulators (TFs) through GH signaling in a sex-differential manner (Waxman et al. 1991, Waxman et al. 1995, Zhang et al. 2012). GH-dependent TFs activate or repress their target genes differently in males and females, propagating sex differences to the entire liver transcriptome (Holloway et al. 2008, Conforto et al. 2012, Conforto et al. 2015). Ablation of circulating GH through the surgical removal of pituitary gland (hypophysectomy; hypox) leads to a near global loss of sex-biased gene expression pattern (Wauthier et al. 2010, Connerney et al. 2017), whereas continuous infusion of GH in male mice, delivered via an osmotic pump, leads to the feminization of liver gene expression (Holloway et al. 2006, Lau-Corona et al. 2017). It is therefore, through the sex-dependent GH secretion patterns, liver sex-biased genes are mainly regulated. Due to its prominent sex-biased GH secretion patterns, the rodent model is ideal to study sex-biased gene expression patterns.

### **1.C GH-dependent TFs**

Sex differences in the temporal plasma GH profiles result in the activation of GH signaling in a dynamic, sex-differential manner (Waxman et al. 1995, Gebert et al. 1999). GH binding to a growth hormone receptor activates JAK2, which in turn phosphorylates the signal transducer and activator of transcription 5 (STAT5) (Brooks and Waters 2010, Sedek et al. 2014). Phosphorylated STAT5 would then translocate to the nucleus and bind to a TTC(C/T)N(G/A)GAA motif to regulate transcription of gene targets (Soldaini et al. 2000). STAT5 activity is pulsatile in male rat liver, as a direct response to GH secretion pulses, while in female rat liver, STAT5 activity show a persistent low level (Choi and Waxman 2000). Between the two members of STAT5 proteins, STAT5B, not STAT5A, mediates most GH signaling in liver (Udy et al. 1997), and is responsible for the loss of ~90% of sex-biased gene expression in STAT5-deficient mouse model (Clodfelter et al. 2006). Sex-biased STAT5 binding sites were shown to be enriched for correspondingly sex-biased genes, confirming the action of STAT5 to positively regulate gene transcription (Zhang et al. 2012). STAT5 binding sites are proximal to 35% of sexbiased genes, suggesting a substantial fraction of sex-biased genes are regulated by STAT5, but also leaving the possibility open for other GH-dependent factors in the regulation of sex-bias genes (Zhang et al. 2012). The male-biased transcriptional repressor BCL6 was shown to preferentially bind to female-biased STAT5 target genes in male liver, indicating BCL6 competition for STAT5 binding sites as a mechanism to repress female-biased genes in male liver (Zhang et al. 2012). A similar negative regulation is carried out in female mouse liver by the female-biased transcriptional

repressor CUX2, which was shown to repress male-biased genes (Conforto et al. 2012). In contrast, binding sites for CUX2 overlap female-biased binding sites for HNF6 at a subset of female-biased genes, indicating a role for CUX2 to indirectly activate femalebiased genes. Overall, HNF6 mostly show positive regulatory role as shown by its malebiased binding sites that are proximal to male-biased genes in male liver (Conforto et al. 2015). Taken together, a network of GH-dependent TFs is essential to maintain sexbiased gene expression in mouse liver.

#### **1.D** Classification of sex-biased genes in response to hypophysectomy

Surgical removal of the pituitary gland (hypophysectomy; hypox) ablates all circulating hormone, including pituitary GH, which is used as a model to tease out GH-dependent gene regulations. There are two major classes of liver sex-biased genes as defined by their sex-dependent positive or negative responsiveness to hypox. Class 1 genes are activated by the GH secretion pattern in the sex where they are highly expressed, whereas class 2 genes are repressed by the GH secretion pattern in the sex where they are less highly expressed (Fig. S4-1) (Wauthier et al. 2010, Connerney et al. 2017). Consequently, class 1 genes are down regulated and class 2 are de-repressed (up regulated) after hypox.

# 1.E GH-dependent sex differences in liver epigenome

GH-dependent sex differences are also apparent in the mouse liver epigenome, including at a subset of transcriptional regulatory sites, as defined by open chromatin regions based on DNaseI hypersensitive sites (DHS) (Ling et al. 2010). Sex-biased DHS are significantly enriched for nearby sex-biased genes with matching sex-specificity, implying the tight connection between sex-biased DHS with sex-biased gene transcription (Ling et al. 2010). The majority of sex-biased DHS, however, are distal, suggesting long-range interactions are involved in the regulation of liver sex-bias genes. The majority of male-biased DHS are suppressed in response to the feminization of GH secretion pattern via continuous GH infusion, delivered via an osmotic pump, while a subset of female-biased DHS is induced. Even more striking, a subset of male-biased DHS showed dynamic opening and closing in response to GH pulses (Connerney et al. 2017).

Analysis of male and female chromatin state maps, based on six chromatin marks and DHS, identified the use of H3K27me3 repressive mark to down regulate highly femalespecific genes in male liver (Sugathan and Waxman 2013). Sex-differential chromatin states also characterize sex-biased DHS, not sex-biased genes, highlighting the importance of distal regulatory sites. Integration of chromatin state maps and genome-wide binding data for five GH-regulated TFs highlight the regulatory role for sex-differential chromatin states, including to facilitate sex-dependent chromatin opening at male-biased DHS and sex-dependent transcription activation through sex-biased STAT5 bindings (Sugathan and Waxman 2013). All of these results suggest an epigenetic-driven mechanism plays a major role in the maintenance of sex-specific gene expression in mouse liver. The mechanisms in which sex-differential chromatin states are established is, however, poorly understood. Long noncoding RNAs were shown to interact with chromatin modifying enzyme to regulate gene transcription (Tsai et al. 2010, Guttman et al. 2011, Engreitz et al. 2013). We hypothesize that sex-biased chromatin state regulations are mediated by sex-specific lncRNAs, which prompted us to characterize liver-expressed lncRNAs in Chapter 2.

### 1.F Sex-biased genes during mouse liver development

Widespread sex differences in mouse liver transcriptome are not seen until puberty (Conforto and Waxman 2012, Lowe et al. 2015), where at 4 weeks of age, 13% of adult sex-biased genes showed significant sex-biased expression, as compared to 85% of adult sex-biased genes at 8 weeks of age (Conforto and Waxman 2012). The majority of sex-biased genes in fetal liver, however, are retained in adulthood (Lowe et al. 2015). Male liver showed extensive changes during liver development, where 76% of male-biased genes were up regulated and 47% of female-biased genes were down regulated by 8 weeks of age, as compared to 33% sex-specific genes that are regulated in female liver (Conforto and Waxman 2012). Sex-specific genes showing early sex-biased expression, at 4 weeks of age, include transcriptional repressor CUX2, highlighting the need to analyze sex-biased expression during liver development (Conforto and Waxman 2012).

# Chapter 2 Hepatic lincRNAs: high promoter conservation and dynamic, sexdependent transcriptional regulation by growth hormone<sup>1</sup>

### 2.A Abstract

Long intergenic non-coding RNAs (lincRNAs) are increasingly recognized as key chromatin regulators, yet few studies have characterized lincRNAs in a single tissue across diverse conditions. Here, we analyzed 45 mouse liver RNA-seq datasets collected under diverse conditions to systematically characterize 4,961 liver lincRNAs, 59% of them are novel, with regards to gene structures, species conservation, chromatin accessibility, transcription factor binding, and epigenetic states. To investigate potential for functionality, we focused on the responses of the liver lincRNAs to growth hormone stimulation, which imparts clinically relevant sex differences to hepatic metabolism and liver disease susceptibility. Sex-biased expression characterized 247 liver lincRNAs, with many being nuclear-enriched and regulated by growth hormone. The sex-biased lincRNA genes are enriched for nearby, and correspondingly sex-biased, accessible chromatin regions, as well as sex-biased binding sites for growth hormone-regulated transcriptional activators (STAT5, HNF6, FOXA1, FOXA2), and transcriptional repressors (CUX2, BCL6). Repression of female-specific lincRNAs in male liver, but not that of male-

<sup>&</sup>lt;sup>1</sup> A modified version of this chapter was published in Molecular and Cellular Biology. Melia, T., Hao, P., Yilmaz, F., and D.J. Waxman (2015). "Hepatic long intergenic noncoding RNAs: high promoter conservation and dynamic, sex-dependent transcriptional regulation by growth hormone." *Mol Cell Biol* **36**:50. qPCR on somatostatin-deficient and STAT5a/STAT5b-KO mice were carried out by Feyza Yilmaz, as noted in appropriate places in this chapter. Dr. Pengying Hao participated in early stage experimental and computational efforts to identify sex-specific lincRNAs.

specific lincRNAs in female liver, was associated with enrichment of H3K27me3associated inactive and poised enhancer states. Strikingly, we find that liver-expressed lincRNA promoters are more highly species-conserved and have a significantly higher frequency of proximal binding by liver transcription factors than liver-expressed proteincoding promoters. Orthologs in one or more supraprimates were identified for many liver lincRNAs, including two rat lincRNAs that show the same growth hormone-regulated, sex-biased expression as their mouse counterparts. The integrated analysis of liver lincRNA chromatin states, transcription factor occupancy and growth hormone regulation provides novel insights into the expression of sex-specific lincRNAs and their potential for regulation of sex-differences in liver physiology and disease.

#### **2.B Introduction**

High-throughput sequencing of mammalian transcriptomes has revealed near-ubiquitous transcription of the genome and the generation of large numbers of non-coding (nc) transcripts. ncRNAs have drawn much attention as potential chromatin regulators, exemplified by classical ncRNAs, such as Xist (Yang et al. 2014). Several thousand ncRNAs have been discovered in human (Cabili et al. 2011, Derrien et al. 2012), mouse (Guttman et al. 2009, Guttman et al. 2010, Kutter et al. 2012, Luo et al. 2013), zebrafish (Ulitsky et al. 2011, Kaushik et al. 2013), and fruit fly (Li et al. 2009, Young et al. 2012), where they exert diverse regulatory roles in gene expression (Rinn et al. 2007, Huarte et al. 2010, Guttman et al. 2011, Ulitsky et al. 2011, Sauvageau et al. 2013); however, the vast majority of ncRNAs are poorly characterized, both computationally and

experimentally. Many ncRNAs share salient features of protein-coding genes, including transcription by RNA polymerase II, 5' capping, splicing, polyadenylation, and deposition of histone marks associated with transcription, specifically H3K4me3 at the promoter and H3K36me3 across the gene body (Guttman et al. 2009). These ncRNAs are typically >200 nt in length (long ncRNAs) and can be classified based on their genomic location in relation to the nearest protein-coding genes, as antisense, intronic, divergent, and intergenic (Rinn and Chang 2012, Kung et al. 2013).

Long ncRNAs have diverse functions, which can be broadly categorized as scaffolds/guides, decoys, enhancers, and others (Rinn and Chang 2012). Scaffold (guide) long ncRNAs bind one or more proteins to form complexes that can direct ncRNA binding partners to target genes. Examples include long ncRNA scaffolds that bind chromatin-modifying complexes capable of reading, writing, and/or erasing histone modifications, indicating a role for long ncRNAs in epigenetic control (Khalil et al. 2009, Tsai et al. 2010, Zhao et al. 2010, Guttman et al. 2011, Wang et al. 2011). HOTAIR, a 2.2 kb long intergenic ncRNA (lincRNA) encoded within the *HOXC* gene cluster, regulates its distal target genes *via* two modular binding domains: one domain binds PRC2, which deposits H3K27me3 (repressive) marks, and a second domain binds LSD1, which demethylates the active chromatin mark H3K4me2 (Tsai et al. 2010). Decoy long ncRNAs, such as Gas5, contain binding motifs that titrate DNA binding proteins. Gas5, which contains a glucocorticoid response element motif embedded within a hairpin structure, is induced during starvation, enabling it to sequester glucocorticoid receptor

and thereby suppress the expression of metabolic genes (Kino et al. 2010). Other long ncRNAs have enhancer-like functions and can activate the transcription of genes either in *cis* or *trans* (Orom et al. 2010, Engreitz et al. 2013, Hacisuleyman et al. 2014). In one study, 7 of 12 long ncRNAs examined exerted positive transcriptional effects within 300 bp of their sites of transcription (Orom et al. 2010). Other long noncoding RNAs bind microRNAs (Wang et al. 2010, Braconi et al. 2011, Ulitsky et al. 2011), regulate alternative splicing (Tripathi et al. 2010), or pair with mRNAs *via* Alu repeats (Gong and Maquat 2011). Thus, long ncRNAs constitute a rich regulatory layer that can shape gene expression by diverse mechanisms.

Mammalian liver plays a critical role in the metabolism of structurally diverse drugs, steroids, fatty acids and environmental chemicals (Waxman and Holloway 2009, Baik et al. 2011, Gu and Manautou 2012). The expression of genes that control liver metabolism and other functions can be dramatically altered by diverse stimuli, making the liver an ideal model for studies of condition-specific gene regulation and for investigation of ncRNAs that may contribute to these processes. Recent studies report the discovery of 10,571 mouse lincRNAs across six tissues, including liver (Luo et al. 2013). Examination of liver lincRNA transcription evolution in three rodent models revealed that nearly half of 388 identified lincRNAs were acquired or lost since the divergence of rat and mouse (Kutter et al. 2012). Another study investigated long ncRNA expression during liver maturation and identified 433 pairs of long ncRNAs and their neighboring protein-coding genes, which showed correlated patterns of expression during neonatal, adolescent and

adult stages of liver development (Peng et al. 2014), indicating *cis* regulation. Despite these efforts, large numbers of liver-expressed long ncRNAs are uncharacterized, and their expression dynamics under conditions that affect liver physiology and disease, including diverse hormonal environments, remain unknown. Comprehensive analyses that connect the responses of lincRNAs to diverse stimuli to changes in chromatin environment, transcription factor (TF) binding, and gene expression are needed to understand liver lincRNA function within the context of gene regulatory networks.

Liver gene expression is highly responsive to many drugs and environmental chemicals (Amacher 2010, Zanger and Schwab 2013), which activate nuclear receptors, such as CAR and PXR, master regulators of hepatic drug and fatty acid metabolism (Tojima et al. 2012, Kodama and Negishi 2013, Banerjee et al. 2014). Hormonal factors also induce large changes in gene expression in the liver (Waxman and O'Connor 2006, Ruiz et al. 2013), in particular genes showing sex-differential expression (Lichanska and Waters 2008, Waxman and Holloway 2009), as is also seen in several other non-reproductive tissues (Rinn and Snyder 2005, Isensee et al. 2008, Sakakibara et al. 2013). In liver, sexbiased gene expression is regulated by the sex-dependent patterns of pituitary growth hormone (GH) release (Waxman and O'Connor 2006): pulsatile GH release in males versus persistent (near continuous) GH release in females (Jansson et al. 1985, Veldhuis et al. 2001, Adams et al. 2015). The resulting sex differences in plasma GH profiles activate the TF STAT5 (Kang et al. 2013) by JAK2-catalyzed tyrosine phosphorylation (Brooks et al. 2014) in a dynamic, sex-differential manner (Waxman et al. 1995, Gebert

et al. 1999), which is linked to downstream sex differences in liver chromatin accessibility and epigenetic (histone) marks (Ling et al. 2010, Sugathan and Waxman 2013). These sex-differential chromatin states are associated with sex differences in the binding of STAT5 and other GH-regulated TFs (Conforto et al. 2012, Zhang et al. 2012, Conforto et al. 2015), which together regulate the sex-biased transcription of ~1,000 liver-expressed genes (Clodfelter et al. 2006, Wauthier and Waxman 2008, Wauthier et al. 2010). Little is known, however, about the mechanisms whereby GH establishes and maintains the unique chromatin states seen in male and female liver (Sugathan and Waxman 2013).

Sex-biased liver gene expression, as seen in mice, rats and humans, is associated with hormone-dependent sex differences in susceptibility to fatty liver disease (Ayonrinde et al. 2011), hepatocellular carcinoma (Ruggieri et al. 2010, Baik et al. 2011), polygenic dyslipidemia, and coronary artery disease (Zhang et al. 2011, Papp et al. 2012). Previous studies identified several lincRNAs, including H19 (Matouk et al. 2007, Zhang et al. 2013), Meg3 (Braconi et al. 2011, Anwar et al. 2012) and HULC (Panzitt et al. 2007, Wang et al. 2010), which are associated with diseases showing a significant sex-biased prevalence in humans, notably hepatocellular carcinoma, liver fibrosis and liver regeneration (He et al. 2014, Takahashi et al. 2014). However, it is unknown whether these or other lincRNAs are expressed in a sex-dependent manner in liver, or whether they might be subject to tight regulation by GH and have the potential to contribute to the widespread sex differences that characterize liver function and disease susceptibility.

Here, we use the mouse model to elucidate gene structures, species conservation, and expression patterns for 4,961 liver-expressed lincRNAs. Strikingly, we find that liver lincRNAs show exceptional per-base conservation and a significantly higher frequency of TF occupancy in their proximal promoter regions than promoters of liver-expressed RefSeq protein-coding genes. We identify a subset of liver lincRNAs showing strong sex-biased expression regulated by pituitary GH secretion profiles, with enrichment for nearby sex-dependent Dnase I hypersensitive sites (DHS) and sex-differential binding by GH-regulated TFs. These findings are integrated with chromatin state analysis, which reveals distinct roles for sex-differential patterns of chromatin states in regulating malespecific compared to female-specific lincRNAs. These findings are discussed in the context of mechanisms of GH regulation of sex-specific liver lincRNAs and their potential for regulation of sex-differences in hepatic physiology and disease.

## 2.C Materials and methods

**Mouse treatment and liver RNA isolation.** Intact and hypophysectomized adult male and female CD1 mice (8-12 weeks old) were purchased from Charles River Laboratories (Wilmington, MA). Livers from male and female wild type and somatostatin-deficient mice (9-12 week old) were provided by Drs. R. M. Luque and R. D. Kineman (University of Illinois at Chicago, Chicago, IL) and were described previously (Meyer et al. 2009, Cordoba-Chacon et al. 2011). Liver tissues were collected from 8-12 week old male and female hepatocyte-specific STAT5a/STAT5b-KO mice, and corresponding floxed control mice, provided Dr. L. Hennighausen (NIDDK, NIH, Bethesda, MD) (Holloway et al. 2007). RNA was isolated from individual mouse livers using TRIzol reagent (Invitrogen Life Technologies, Inc., Carlsbad, CA). Animals were treated using protocols specifically reviewed for ethics and approved by the Boston University Animal Care and Use Committee.

Mouse liver RNA-Seq and data analysis. RNA-seq data was obtained from 45 poly(A)selected mouse liver RNA-Seq samples representing five biological conditions. Individual RNA samples were prepared from either total liver RNA (22 samples) or nuclear liver RNA (23 samples). The biological conditions represented in the total liver RNA datasets are untreated male and untreated female mouse liver, and livers of male mice treated with chemical activators of the nuclear receptors CAR, PXR and PPARA. Biological conditions represented in the nuclear RNA datasets are untreated male and female mouse liver, and livers of male mice treated with chemical activators of the nuclear receptors CAR and PXR. A detailed analysis of the lincRNAs responsive to chemical activators of the nuclear receptors will be presented elsewhere. Liver RNA samples were divided into 8 sets, each comprised of RNA samples sequenced in the same run (Table S7 in (Melia et al. 2016), Gene Expression Omnibus (GEO) series GSE66140 and GSE48109, and ENCODE series GSE36025 and GSE6026). All datasets were mapped to the mouse genome (mm9) using TopHat2 (Kim et al. 2013) with default parameters. A mouse genome mm9 annotation, derived from the UCSC repository, was downloaded from Illumina (archive-2012-03-09-05-07-56) and was used as the starting

point for Tophat2 to discover splice junctions. For paired-end reads, the *r* parameters were calculated by subtracting the length of the adaptor sequence (2 x 120 nt) and read length from the average fragment size. The fragment size was determined experimentally from Agilent Bioanalyzer tracings or was determined computationally by mapping paired-end reads to the mm9 genome using Bowtie (Langmead et al. 2009) and then computing the mean distance between every mapped paired read. Mapping percentages varied from 68% (ENCODE samples) to 98%, with a median of 85% reads mapped.

**Defining lincRNAs.** All 45 RNA-Seq samples were used for transcript assembly by Cufflinks (Trapnell et al. 2010), which finds the minimum number of transcripts that encompass all the reads while including each read in at least one transcript. The mouse genome mm9 annotation was used as the starting point for transcript reconstruction. Fragment bias correction and multi-read correction options were both applied; all other arguments were set to default values. A total of 31,792 transcribed loci (transcribed genes), 120,376 isoforms, and 63,656 splicing variants were identified. The initial list of 31,792 transcribed genes was reduced to 7,088 transcribed loci after filtering to remove transcripts that overlap RefSeq protein-coding genes in either the sense or antisense direction. We next applied three filters to remove transcripts that are lowly expressed (low expression filter, described below), short (RNA length < 200 nt), or exhibit similarity to protein-coding genes (protein-coding potential filter; described below), reducing the number of transcripts to 4,961 (liver-expressed lincRNAs). To obtain a stringent set of liver-expressed lincRNAs, we further filtered the list of 4,961 by removing: 1) lincRNAs that overlap with any of the 1,626 miRNAs, 334 rRNAs, 433 tRNAs, 1,424 snRNAs or 10,517 pseudogenes annotated in Ensembl version NCBIM37.67, in either the sense or antisense direction; and 2) lincRNAs whose longest ORF was > 150 amino acids, resulting in a set of 4,454 stringent lincRNAs (Fig. 2-1A).

Low expression filter. The quality of transcript reconstruction results depends on the depth of sequencing, and to a certain extent, the level of gene expression; genes that are highly expressed are more likely to have reads covering their full gene bodies than lowly expressed genes. Accordingly, we excluded lowly expressed genes, as their gene boundaries definitions are likely truncated. We quantified the expression of each gene (in FPKM) by rerunning Cufflinks with the –G option turned on. To determine the proper gene expression cutoff for each RNA-Seq sample, we varied the expression cutoff (in FPKM) while counting the fraction of RefSeq noncoding gene ends that are correctly predicted (i.e. covered by reads from the 5' to the 3' end) versus those are incorrectly predicted for each cutoff value. We plotted a receiver operating characteristic (ROC) curve for each RNA-Seq sample based on the number of correct and incorrect predictions for each cutoff interrogated, and then selected the cutoff with the greatest true positive rate while keeping the false positive rate below 15%. We thus identified optimized FPKM cutoffs tailored to the sequencing depth and read length of each RNA-Seq sample. Genes that passed the expression cutoff in at least two biological replicates in at least one dataset were retained. For the G81 dataset (Table S7 in (Melia et al. 2016)), we retained genes that passed the expression cutoff in at least two biological conditions. Expression

cutoff values ranged from 0.01 to 0.09 FPKM for 35 of the 45 RNA-seq samples, and from 0.6 to 1.4 FPKM for the other 10 samples (G78, and ENCODE S1, S2 samples; Table S7 in (Melia et al. 2016)).

**Protein-coding filter**. Transcripts that contain a known protein domain, or do not exhibit codon degeneracy when looking at their multiple genome alignments, were identified (see below) and removed. To identify known protein domains, we created mature RNA sequences for each isoform of every gene by downloading their genomic sequences, removing introns and then concatenating consecutive exonic sequences. Bases that were not assigned to any of the four bases were replaced with either A, C, T or G with equal probabilities. The resulting mature RNA sequences were then translated to amino acid sequences in all 6 reading frames. For each amino acid sequence, we calculated the probability that it belongs to one of the 13,672 protein families defined in Pfam (Finn et al. 2014) using HMMER (Finn et al. 2011), which enabled us to identify, and then filter out, 480 transcribed loci that contain at least one protein domain.

**Codon evolutionary analysis.** Codon evolutionary analysis was used to distinguish coding from noncoding regions based on the observation that protein-coding genes are depleted from non-synonymous and nonsense substitutions during evolution to retain their amino acid sequence (Lin et al. 2011). Briefly, based on the assumption that each codon evolved independently as a Markov process along a phylogenetic tree, we calculated the probability for a set of codons to evolve from a phylogenetic tree that

represents a coding region compared to another phylogenetic tree for a noncoding region. Both trees are identical in terms of branch lengths and structures; the difference lies in the frequency of a specific codon evolving to another codon (i.e. codon substitution matrix). The rate of codon substitution among identical and synonymous codons is expected to be higher for the coding compared to noncoding region. We downloaded 64 x 64 codon substitution matrices for coding regions and non-coding regions along with a phylogenetic tree of 29 mammals from Lindblad-Toh et al (Lindblad-Toh et al. 2011), which were empirically derived from sequence alignments of many known coding and noncoding sequences and maximized using an expectation-maximization approach. We used PhyloCSF (Lin et al. 2011) with these downloaded parameters to calculate the probability of each gene being coding or noncoding, as follows. For each gene, we extracted 46 species whole-genome multiple alignment data from the UCSC genome browser sequence and annotation downloads server (Karolchik et al. 2014), removed species that were not in the defined set of 29 mammals, enumerated all possible complete start-to-stop codon regions in the 6 possible reading frames, calculated the log-likelihood ratio based on the coding model versus the noncoding model for each reading frame and reported the highest score. The resulting log-likelihood score is termed coding potential score, where a positive score indicates that the gene being interrogated is likely to encode conserved peptides. We identified 126 genes with positive coding potential out of 4,252 putative lincRNA genes that could be multi-genome aligned. 4,961 lincRNA genes remained after applying the above three filters (Fig. 2-1A).
Other high-throughput datasets. We downloaded genome-wide chromatin mark and transcription factor binding data sets for mouse liver, as follows: H3K4me3 marks from Sugathan and Waxman (GEO series GSE44571) (Sugathan and Waxman 2013) and Kutter et al (ArrayExpress E-MTAB-867) (Kutter et al. 2012); H3K36me3 from Sugathan and Waxman (GEO series GSE44571) (Sugathan and Waxman 2013) and the mouse ENCODE project (GEO series GSE31039) (Stamatoyannopoulos et al. 2012); RNA polymerase II from Sun et al (GEO series GSE21773) (Sun et al. 2011) and the mouse ENCODE project (GEO series GSE36027) (Stamatoyannopoulos et al. 2012); STAT5 and BCL6 from Zhang et al (GEO series GSE31578; STAT5-high male liver peak set using antibody C-17, Santa Cruz Biotechnology #sc-835x; male and female liver peak sets for BCL6 using antibody N-3, Santa Cruz Biotechnology #sc-858x (Zhang et al. 2012) (Table S11 in (Melia et al. 2016), and see below); HNF6 from Conforto et al (GEO series GSE60014; male liver peak set) (Conforto et al. 2015); CUX2 from Conforto et al (GEO series GSE35985) (Conforto et al. 2012);FOXA1 and FOXA2 from Li et al (ArrayExpress E-MTAB-805; merged male and female liver peak sets) (Li et al. 2012); HNF4A from Schmidt et al (GEO series GSE22078) (Schmidt et al. 2010); PPARA, LXR and RXR from Boergesen et al (GEO series GSE35262) (Boergesen et al. 2012); HNF1A from Faure et al (ArrayExpress E-MTAB-941) (Faure et al. 2012); CEBPA and CEBPB from Jakobsen *et al* (GEO series GSE42321) (Jakobsen et al. 2013); GR from Everett et al (GEO Series GSE45674) (Everett et al. 2013); DHS from Ling et al (GEO series GSE21777; combined set of 70,048 male-specific and sex-independent liver DHS) (Ling et al. 2010); and poly(A) sites from Derti et al (GEO series GSE30198)

(Derti et al. 2012). We also downloaded 2,651,801 multi-tissue, mouse-specific CAGE clusters from the FANTOM consortium (Balwierz et al. 2009).

**Expression quantification.** Gene expression was quantified using HTSeq (Anders et al. 2015) using default settings. The same mouse genome mm9 annotation that was used for mapping reads was used to define RefSeq gene and exon boundaries for quantifying gene expression. To facilitate quantification of gene expression, each gene whose direction of transcription is unknown was assigned a provisional directionality based on a simple majority vote of reads mapping to the plus versus the minus strand across the set of 18 stranded RNA-Seq samples examined. In the case of a tie, the direction of transcription was arbitrarily assigned to the plus strand by default. In this manner, 1,171 lincRNAs were assigned to the plus strand, 1,150 to the minus strand, and 67 lincRNAs defaulted to the plus strand. The geometric mean normalization method (Anders and Huber 2010) was used to normalize reads across samples for most analysis. For expression distribution analysis, reads were normalized across samples by calculating FPKM values. The prcomp function in R (Gentleman et al. 2004) was used to compute principal components of read counts across samples that were normalized with a variance stabilizing transformation (Anders and Huber 2010) to remove the dependence between mean and variance of count data. The set of 500 lincRNAs with the highest gene expression variance across the 45 samples was used to calculate the principal components. We plotted the first against the fifth principal components, as they separate the samples either by their cellular compartment (total vs. nuclear RNA) or sex.

**Subcellular localization.** Differential analysis was carried out using DESeq (Anders and Huber 2010) to compare the normalized gene expression values for nuclear versus total liver RNA samples for the G78 RNA-Seq dataset, and separately, for the G81 RNA-Seq dataset (Table S7 in (Melia et al. 2016)). The two differential analyses were performed separately on liver-expressed lincRNAs and RefSeq protein-coding genes. For each dataset (i.e. G78 and G81), lincRNAs and RefSeq protein-coding genes that have < 50 reads summed up across conditions in all samples were excluded. Subcellular localization was assigned when at least one dataset showed significant enrichment for gene expression in either nuclear or total RNA at fold-change  $\geq$  2 and adjusted p  $\leq$  0.05.

Liver-specificity. 42 paired-end RNA-Seq samples collected from 21 mouse tissues (GSE39524; Table S8 in (Melia et al. 2016)) were downloaded from the mouse ENCODE project (Stamatoyannopoulos et al. 2012). Each sample was mapped to the mouse genome. The expression in liver and in the other 24 non-liver samples was quantified for each of the 4,961 liver-expressed lincRNA and 14,057 liver-expressed RefSeq protein-coding genes (see 'Expression quantification', above). The tissue-specificity of each gene (lincRNA or RefSeq protein-coding gene) was quantified by dividing its expression level in a given tissue over the sum of its expression levels across all 22 tissues. To correct for the over representation of liver samples in the full tissue panel (which includes data for only two samples for all tissues except liver), all liver data are also presented as two samples, one based on the maximum expression of each gene

across 42 total liver RNA samples and the other based on the maximum expression across 23 nuclear liver RNA samples. These 65 liver RNA-seq data sets encompass the 45 samples used to discover lincRNA gene structures, as well as 8 RNA-seq samples from the hypophysectomy study (G85 samples) and 12 RNA-seq samples from the continuous GH treatment study (G88 samples) (Table S1, Table S7 in (Melia et al. 2016); GEO series GSE66140). A gene was considered as solely expressed in a given tissue when no sequence reads were detected in other tissues. Tissue specificity was also determined based on the maximum expression level of each lincRNA in a given tissue, measured in FPKM. The highest FPKM amongst all RNA-seq samples in a given tissue was used to represent the gene expression level for such tissue. RNA-seq data was also downloaded from ArrayExpress E-MTAB-2582 (Stubbington et al. 2015) and mapped as described above (Table S8 in (Melia et al. 2016)) to identify GH-regulated sex-specific liver lincRNAs expressed in three T cell subtypes where STAT5 activity is high.

**Conservation analysis.** We used Phastcons (Siepel et al. 2005) to calculate per base conservation scores based on the multiple genome alignment of 10 selected species from the supraprimate (Euarchontoglires) clade (rat, guinea, rabbit, human, chimp, orangutan, rhesus, marmoset, bushbaby, treeshrew), downloaded from the UCSC genome browser sequence and annotation downloads server (Karolchik et al. 2014). A multi-genome alignment was derived from the 30 vertebrate species multi-genome alignment by removal of the 19 species excluded from our analysis (i.e. retaining only mouse and the other 10 species). In separate analyses, putative orthologs in the same 10 species were

identified based on sequence similarity, as follows. First, we collapsed all isoforms of each mouse lincRNA gene into a single transcript by taking the union of all exons. Second, we downloaded each genome, where all of the repeats are hardmasked (i.e. replaced by N's). Third, we used Exonerate (Slater and Birney 2005) to find sequences in each genome that are similar to each lincRNA with the affine:local model selected and softmasking of lincRNA introns. Up to 20 top hits were retained. Fourth, we concatenated neighboring hits, i.e., hits within 2 kb of each other. Fifth, we realigned the concatenated hits with their orthologous lincRNA sequence using Exonerate with the global alignment option selected. Where no global alignment was produced, we accepted any local alignment of the two sequences. Based on this final realignment, we calculated the percent overall identity of the two sequences. Sixth, we removed any hit that overlapped with an annotated RefSeq protein-coding gene in the genome of interest and retained alignments that spanned at least 10% of the mature mouse lincRNA gene and were > 600 bp in length. These cutoffs were chosen to maximize the frequency of correct prediction of known lincRNA orthologs (Fig. S10A, Fig. S10B in (Melia et al. 2016)). No alignment percent identity cutoff was required, as the lowest percent identity match was ~60% (Fig. S10C in (Melia et al. 2016)). To identify genomic regions in the rat that are syntenic with mouse lincRNAs, we used the LiftOver tool from the UCSC genome browser (Karolchik et al. 2014) by relaxing the identity match to 60% and by not allowing the hit to span multiple orthologous regions.

**qPCR.** Primers used for confirmatory quantitative PCR to assess the sex-specificity and GH-responsiveness of five mouse lincRNAs and two orthologous rat lincRNA regions are listed in Table S10 in (Melia et al. 2016). Real-time PCR analysis was performed on RNA samples isolated from individual mouse and rat livers using Power SYBR green PCR master mix and an ABS 7900HT sequence detection system (both from Applied Biosystems).

**Overlap analysis.** Overlap with a known lincRNA found in lncRNAdb (Amaral et al. 2011) was established by the overlap between the liver-expressed lincRNA or its orthologs, identified as described above, in either the sense or antisense direction. Overlap between a liver-expressed lincRNA and a PRC2-interacting RNA (Zhao et al. 2010) further required that the lincRNA be transcribed in the same direction as the PRC2-interacting RNA, as annotated in the original study.

Comparison of TF occupancies for FPKM-matched lincRNA and protein-coding genes. For each liver-specific lincRNA whose median FPKM across datasets was > 0.52 FPKM, we randomly chose a liver-specific protein-coding gene with a similar FPKM ( $\pm$ 0.1 FPKM). This FPKM-matched lincRNA and protein-coding gene set was used to determine the TF occupancy rate of the promoter region for each gene set. Liver-specific lincRNA and liver-specific protein-coding genes were defined as genes whose expression in liver is  $\geq$  60% of their cumulative expression across 22 mouse tissues. Multi-tissue expressed genes were defined as genes expressed in at least two tissues, with no more than 10% of their cumulative-expression across 22 tissues occurring in liver, and < 60% of their cumulative-expression across 22 tissues occurring in any non-liver tissue.

Sex-specificity and GH regulation of liver lincRNAs. Normalized gene expression values from six sets of male and female mouse liver RNA-Seq samples were compared by differential expression analysis to identify sex-specific lincRNAs, as follows. DESeq and edgeR (Robinson et al. 2010) were used to analyze the total RNA datasets G78 (T1), G83 (T2) and G85 (T3). Normalized gene expression ratios were used for two unreplicated G78 nuclear RNA samples (N1). Multi-factor DESeq was used to account for different sequencing runs for the merged dataset of G83 and G85 total RNA samples (T4), and DESeq was used for the merged dataset of G81 and G86 nuclear RNA samples (N2). Liver-expressed lincRNAs that showed a  $\geq$  4-fold difference in expression between male and female liver at adjusted p < 0.05 in one or more of the 6 datasets, by either DESeq or edgeR, were tentatively designated sex-specific lincRNAs. For the G78 nuclear dataset, a threshold of > 50 read counts, summed across all male and female samples, was used in place of a p-value cutoff. Further, for nuclear RNA datasets, we excluded from the sex-specific list any lincRNA that did not show a response to either hypophysectomy (at fold change  $\geq$  4 and adjusted p  $\leq$  0.05) or continuous GH infusion (at fold change  $\geq$  2 and adjusted p < 0.05), resulting in a list of 247 sex-specific liver-expressed lincRNAs. 172 of the sex-specific lincRNAs were expressed at > 1 FPKM (Table S1 in (Melia et al. 2016)). 171 of the 247 lincRNAs were designated stringent sex-specific lincRNAs, based on their dependence on GH for sex-specific expression, as shown by their dysregulation

in male liver following continuous infusion of GH for 1, 4 or 7 days (female-like GH pattern) or following hypophysectomy, and were used for TF enrichment analysis. 141 of the 171 stringent sex-specific lincRNAs could be classified based on their responsiveness to hypophysectomy (class 1 or class 2 responses; Table 2-2). The GH-regulated sexspecific lincRNAs showed a sex-difference in expression ranging from 4-fold to 190fold, for male-specific lincRNAs, and 4.2-fold to 1,240-fold, for female-specific lincRNAs (see Fig. 2-4B, below). Any lincRNA from the list of 4,961 liver-expressed lincRNAs that was not included in the list of 247 sex-specific lincRNAs was deemed sexindependent; the latter list was used as a background when calculating the enrichment of sex-specific lincRNAs nearby sex-specific protein-coding genes. The 857 sexindependent lincRNAs showing < 1.2-fold sex-difference in at least 3 of 4 datasets (G78 nuclear, G78 total, G83 and G85 datasets) were designated stringent sex-independent lincRNAs. DESeq and edgeR were used to identify liver-expressed lincRNAs showing differential expression between intact and hypophysectomized mice, and those showing differential expression between male mice treated with GH as a continuous infusion for 1, 4 or 7 days compared to vehicle-treated control male mice based on normalized gene expression values. Hypophysectomy and GH infusion both lead to widespread dysregulation of sex-specific protein-coding genes in mouse and rat liver (Wauthier and Waxman 2008, Wauthier et al. 2010). The significance level was set at fold-change > 4for the hypophysectomy model and fold-change > 2 for at least one of the time points in the continuous GH infusion model, and adjusted p < 0.05 for both. RNA-seq data sets are available as GEO series GSE66140.

Sex-specific protein-coding genes. DESeq and edgeR were used to identify a set of 840 sex-specific RefSeq genes based on the normalized gene expression values in the G78, G83 and G85 total RNA datasets, of which 33 are ncRNAs. Thus, 807 protein-coding genes were identified as sex-specific based on the criteria adjusted  $p \le 0.05$  in at least 2 of the 3 total RNA datasets, as determined by either DESeq or edgeR. The 807 RefSeq coding genes showed a sex-difference in expression ranging from 1.41-fold to 884-fold, for male-specific genes, and 1.41-fold to 10,030-fold, for female-specific genes, when considering the highest fold change across the three datasets (see Fig. 2-4B, *below*). Any RefSeq protein-coding gene that was not included in the sex-specific protein-coding genes list was considered as sex-independent.

**Enrichments.** 1) <u>Gene proximity enrichment</u>: Enrichment scores (ES) for the proximity of sex-specific lincRNAs to be nearby (within 25 kb) sex-specific RefSeq protein-coding genes of the same sex specificity were calculated as follows: ES = (number of sex-specific lincRNAs nearby matched sex-specific RefSeq protein-coding genes) / (number of sex-specific lincRNAs nearby sex-independent RefSeq protein-coding genes), as compared to (number of sex-independent lincRNAs nearby any sex-specific RefSeq protein-coding genes) / (number of sex-independent lincRNAs nearby sex-independent RefSeq protein-coding genes) / (number of sex-independent lincRNAs nearby sex-independent RefSeq protein-coding genes) / (number of sex-independent lincRNAs nearby sex-independent RefSeq protein-coding genes) / (number of sex-independent lincRNAs nearby sex-independent RefSeq protein-coding genes). 2) <u>Enrichment of DHS and TF binding events (except for CUX2 binding sites)</u>: ES values for having a female-specific TF binding site or DHS nearby (within 10 kb) a female-specific lincRNA was calculated as follows: ES =

(number of stringent female-specific lincRNAs that have a nearby female-specific DHS, or a nearby stringent female-specific TF binding site) / (number of stringent femalespecific lincRNAs nearby non female-specific (i.e. male-specific + sex-independent) DHS, or TF binding sites), as compared to (number of non female-specific lincRNAs (i.e. male-specific + stringent sex-independent lincRNAs) that have a nearby DHS, or femalespecific TF binding site) to (number of non female-specific lincRNAs nearby non femalespecific DHS, or TF binding sites). Enrichment analysis was carried out in the same manner for a control gene set, comprised of 267 sex-independent lincRNAs that are hypophysectomy-responsive (see Fig. 2-4C, *below*), which were compared to the set of sex-independent lincRNAs that are not responsive to hypophysectomy. A corresponding analysis was performed for male-specific DHS and TF binding site enrichments. 3) CUX2 binding enrichment: Enrichment scores for having CUX2 bound within 10 kb of a stringent female-specific lincRNA were calculated as follows: ES = (number of stringentfemale-specific lincRNAs that have a nearby CUX2 binding site) / (number of stringent female-specific lincRNAs that do not have a nearby CUX2 binding site), as compared to (number of non female-specific lincRNAs (i.e. male-specific + stringent sex-independent lincRNAs) that have a nearby CUX2 binding site) / (number of non female-specific lincRNAs that do not have a nearby CUX2 binding site). Since CUX2 is not expressed in male liver, no enrichment analysis was carried out for male-specific CUX2 binding. 4) Gene set enrichment analysis (GSEA) (Subramanian et al. 2005) was carried out using default settings to determine if sex-specific lincRNA TSS regions (-1 kb to +1 kb window surrounding the TSS), or gene bodies (TSS to TES) are enriched for having a

particular chromatin state in one sex, but not the other. These analyses using chromatin state maps in male and female mouse liver (Sugathan and Waxman 2013), where each consecutive 200 bp segment of the mouse genome was assigned one of 14 chromatin states (state 1 to state 14; see Fig. 2-6B, *below*). Enrichment for chromatin state *x* in male liver was established by comparing the distribution of liver lincRNAs where state *x* is found within its gene body or TSS in male liver, but not in female liver, over the full list of 4,961 liver lincRNAs, ranked from high to low male/female expression ratio. This analysis was carried out for each of the 14 states (Sugathan and Waxman 2013) in male liver, and separately, in female liver. Normalized enrichment scores  $\geq$  1.45 with FDR  $\leq$ 0.05 were deemed significant. GSEA was also carried out for 16,124 RefSeq proteincoding genes, where we excluded genes that have no reads in the RNA-seq datasets used to determine sex-specificity.

**Clustering of sex-specific lincRNAs.** Consensus clustering (Reich et al. 2006) was used to identify 5 clusters of male-specific lincRNAs based on the chromatin state patterns of each lincRNA in a 2 kb window surrounding the TSS, and in a 2 kb window surrounding the TES, in both male and female liver. The number of clusters, n=5, was chosen to give robust, reproducible clusters over 100 runs. Each lincRNA was clustered after counting the total number of bases across the 2 kb region assigned to each of the 14 states for each genomic segment of interest, i.e. male TSS  $\pm$  1 kb, female TSS  $\pm$  1 kb, male TES  $\pm$  1 kb and female TES  $\pm$  1 kb. The 5 clusters were then sorted from high to low male-specificity (clusters A to E). Similar clustering was performed for female-specific lincRNAs. Other analyses. For TF peak analysis, HNF1A and HNF4A raw data were mapped to mm9 using Bowtie2 (Langmead et al. 2009). Biological replicates were combined before removing PCR duplicates using Picard (http://broadinstitute.github.io/picard). Peaks were then discovered using MACS2 (Zhang et al. 2008). BCL6 peaks (binding sites) identified separately in male and female liver (Zhang et al. 2012) were merged to give a single list of 6,656 peaks. This merged peak list was compared to the set of genomic regions differentially bound by BCL6 between individual male and female liver samples (Zhang et al. 2012), as determined by Andy Rampersaud of this laboratory using diffReps (Shen et al. 2013) after removing differential binding regions showing < 2-fold difference between male and female samples, as well as those diffReps regions mapping to weak BCL6 peaks (MACS2 score < 25, or peaks having < 20 ChIP-seq reads). The 6.656 merged BCL6 peaks were thus classified into 1,459 male-biased, 42 female-biased and 5,155 sex-independent BCL6 peaks (Table S11 in (Melia et al. 2016)). Statistics for human lncRNAs were calculated from GENCODE annotation version 21 (June 2014 freeze, GRCh38, Ensembl77). For hypothesis testing, p-values were calculated using R. P-values < 2.2e-16 were rounded up to 1e-15 due to R's limitation for representing floating numbers. All analyses were done using custom Perl and R scripts unless otherwise noted

## Availability of supporting data

The sequencing data sets supporting the results of this article are available at the Gene Expression Omnibus (GEO) website (<u>http://www.ncbi.nlm.nih.gov/gds</u>) as series GSE48109 and GSE66140.

## **2.D Results**

**Identification of liver-expressed lincRNAs.** 45 RNA-Seq datasets, obtained using polyA-selected total and nuclear liver RNA isolated from male and female mice under various conditions (see Methods), were used to assemble liver-expressed RNA transcripts and discover novel gene structures. 7,088 transcribed intergenic loci were identified after removing transcripts that overlap RefSeq protein-coding genes on either strand. Transcripts showing low expression across samples were filtered out (Fig. 2-1A). A second filter removed transcripts < 200 nt in length. A final filter removed transcripts whose protein-coding potential score, which indicates the rate of synonymous vs. nonsynonymous substitutions across species, resembles that of protein-coding genes. This final filter applied a stricter cutoff for removing transcripts with protein-coding potential than that used to define RefSeq noncoding genes (Fig. 2-1B). A total of 4,961 liverexpressed lincRNAs passed all three filters (Table S1 in (Melia et al. 2016)). Removal of 188 lincRNAs that overlap small RNAs or pseudogenes on either strand, as well as 319 lincRNAs with an open reading frame (ORF) > 150 amino acids in length, resulted in 4,454 stringent lincRNAs. 92% (4,554) of the full set of liver lincRNAs do not overlap RefSeq transcripts (i.e., non-RefSeq lincRNAs), and 59% (2,941) do not overlap any

ncRNAs reported previously (Guttman et al. 2009, Kutter et al. 2012, Luo et al. 2013, Xie et al. 2014) (Table S1 in (Melia et al. 2016)).

Liver lincRNA gene structures. To distinguish genuine lincRNAs from spurious products of leaky transcription, we looked for associations with histone marks classically found at lincRNA promoters (H3K4me3) and gene bodies (H3K36me3) (Guttman et al. 2009). We examined the frequency of each histone mark at lincRNA genes as compared to protein-coding genes after binning both gene sets based on their expression levels, to remove bias against genes expressed at a low level. The frequency of each mark increased with increasing gene expression in a similar manner for both lincRNA and protein-coding genes (Fig. S1A, S1B in (Melia et al. 2016)), as expected for genuine transcripts. The lower frequencies of these marks at lower expression levels may reflect histone mark intensities below the threshold of detection at the available ChIP-seq sequencing depths, as well as incomplete lincRNA gene structures, which can lead to incorrect assignment of gene body boundaries.

The completeness of our reconstructed gene structures was assessed by proximity of lincRNA transcription start sites (TSS) to 5' cap site locations identified as CAGE-Seq clusters in liver (Balwierz et al. 2009), and by the proximity of lincRNA transcript end sites (TES) to 3' poly(A) tails identified by genome-wide sequencing (Derti et al. 2012). Both marks increased in frequency as the level of lincRNA expression increased, similar to the trend for protein-coding genes (Figs. S1C, S1D in (Melia et al. 2016)). Overall,

71% of liver lincRNAs have either a proximal CAGE-Seq, poly(A)-Seq or promoterassociated H3K4me3 peak, supporting our lincRNA gene boundary assignments and our conclusion that the liver lincRNAs are independent transcripts, and not extensions of nearby protein-coding transcripts. Further, the fraction of lincRNA promoters vs. proteincoding gene promoters bound by RNA polymerase II was very similar at equivalent levels of gene expression (Fig. S1E in (Melia et al. 2016)), suggesting that RNA polymerase II makes a substantial contribution to liver lincRNA transcription.

24% of the liver lincRNAs are multi-exonic (Table 2-1, Table S1 in (Melia et al. 2016)). This compares to only 10-15% multi-exonic lincRNAs in ribosomal RNA-depleted mouse and rat liver RNA-seq libraries (Kutter et al. 2012). The multi-exonic liver lincRNAs tend to have longer exons and longer mature transcripts, but intron lengths similar to protein-coding genes (Table 2-1).

Liver lincRNAs: nuclear enrichment and tissue specificity. The 4,554 non-RefSeq liver lincRNAs are expressed at a lower level than RefSeq protein-coding genes, similar level to known RefSeq noncoding genes, and were most highly expressed in the nuclear compartment (Fig. 2-1C), indicating nuclear retention. Of 2,399 liver lincRNAs having sufficient sequencing depth for analysis, 1,234 showed significant nuclear enrichment, while only 93 were enriched in the total RNA fraction (primarily cytoplasmic RNA) (Table S1 in (Melia et al. 2016)). In contrast, RefSeq protein-coding transcripts showed almost equal numbers of nuclear-enriched as total RNA-enriched transcripts (33% vs.

29% of transcripts). Subcellular compartmentalization captured ~70% of the variance in lincRNA expression across the liver RNA datasets and cleanly separated total liver RNA from nuclear liver RNA samples (Fig. 2-1D, *left*). A more minor component (~3% of the variance) separated the liver lincRNAs based on sex of the individual mice (Fig. 2-1D, *right*) (also see below).

27% of the 4,961 liver lincRNAs showed strong liver-specificity, defined as  $\geq$  60% of total expression level summed up across the 22 mouse tissues examined occurring in liver (Fig. S2A, Table S1 in (Melia et al. 2016)), as compared to only 6% of liver-expressed protein-coding genes (Fig. S2B in (Melia et al. 2016)). 1,437 (66%) of the 2,171 lincRNAs expressed in liver at  $\geq$  1 FPKM (fragments per kb length per million sequence reads), corresponding to an estimated one molecule per cell (Mortazavi et al. 2008), were below that minimum expression level in all 21 other tissues examined (Fig. 2-1E, Fig. S3A-S3B, Table S1 in (Melia et al. 2016)). Subsets of liver lincRNAs showed co-expression largely limited to testis or brain (Fig. S3C, S3D in (Melia et al. 2016)). 30 lincRNAs were expressed at  $\geq$  1 FPKM in all 22 tissues (Fig. 2-1E, Table S2 in (Melia et al. 2016)). Only 29 liver lincRNAs were largely invariant in expression across tissues, as compared to 1,069 liver-expressed protein-coding genes (<1.5-fold range of expression across 22 tissues).

**Conservation of liver lincRNAs.** LincRNA conservation was assessed using two complementary approaches: 1) by quantifying per-base conservation along lincRNA

transcripts and regions upstream of their TSS based on a multi-genome alignment of 10 supraprimate species; and 2) by using sequence similarity to identify putative orthologs in the same 10 species. The first approach relies on the accuracy of the multi-genome alignment and the assignment of synteny across species to quantify conservation, whereas the second approach identifies genomic regions in each species most similar in sequence to each mouse lincRNA, independent of their placement in the multi-genome alignment.

To implement the first approach, we calculated the probability of each base to be conserved in a given multi-genome alignment, and then compared the cumulative frequency distribution of the per-base conservation score for three gene-centric regions – exons, introns, and promoters - between lincRNA and RefSeq protein-coding genes. Protein-coding exons showed significantly higher conservation than lincRNA exons. which in turn showed modest but significant conservation compared to a random (background) sampling of untranscribed intergenic regions, indicating the lincRNA exons are subject to moderate evolutionary constraint (Fig. 2-2A). LincRNA introns and RefSeq protein-coding introns both showed very low per-base conservation. In contrast, lincRNA promoters were significantly more conserved than protein-coding promoters (p < 1e-15, Kolmogorov-Smirnov). This pattern of increased lincRNA promoter conservation was even more apparent when comparing lincRNA vs. protein-coding promoters of genes that are liver-specific (Fig. 2-2B, top; dashed vs. solid blue lines), or genes that are multi tissue-expressed (Fig. 2-2B, bottom). Additionally, multi tissue-specific lincRNA promoters showed higher conservation than liver-specific lincRNA promoters (Fig. 2-2B,

*bottom* vs. *top;* compare two sets of blue dashed lines). The strong per-base conservation of lincRNA vs. protein-coding promoters drops abruptly downstream of the TSS; this indicates strong purifying selection operating on lincRNA promoters (Fig. 2-2C, *left*). No such conservation was seen surrounding lincRNA TES (Fig. 2-2C, *right*). Fig. S4C-S4E in (Melia et al. 2016) shows examples of individual lincRNA promoters with high per-base conservation.

Next, we examined whether the greater conservation of lincRNA promoters is associated with increased binding of liver TFs. Indeed, for 11 out of 12 liver TFs examined, liver-specific lincRNA proximal (1 kb upstream) promoters were bound significantly more frequently than an FPKM-matched set of liver-specific protein-coding gene promoters (Fig. 2-2D, *left*; p < 0.05 for each TF, Binomial exact test). Further, 47% of the liver-specific lincRNA promoters are occupied by 3 or more of the 12 liver TFs examined, compared to only 29% for the corresponding protein-coding genes. Promoter region occupancy was more similar when the upstream region was expanded to 25 kb (Fig. 2-2D, *right*), with 72% occupancy by 3 or more factors for lincRNAs vs. 62% for protein-coding genes, consistent with the rapid drop off in lincRNA sequence conservation in the extended upstream region (Fig. 2-2C).

Our second approach to defining lincRNA conservation identified putative orthologs in other species based on sequence similarity, as detailed in Methods. We identified at least one aligned region/putative ortholog for 70% of the liver lincRNAs (Fig. 2-3A), with the

largest number found in rat (60%; 2,984 putative orthologs) (Table S3, Table S4 in (Melia et al. 2016)). 91% (2,730) of the putative rat lincRNAs orthologs are at syntenic regions based on mouse and rat whole genome alignments, indicating that the homology extends to the surrounding genomic regions. Using this approach, we correctly identified 8 out of 12 annotated lincRNA orthologs (Amaral et al. 2011) in rat, human, chimp and/or rhesus (Table S4 in (Melia et al. 2016)). Six of these 8 lincRNAs bind chromatin-modifying enzymes (Table S5 in (Melia et al. 2016)), consistent with reports that many lincRNAs bind chromatin remodelers and contribute to chromatin state regulation (Khalil et al. 2009, Guttman et al. 2011). 350 of the liver lincRNAs are in the set of 9,788 PRC2-interacting RNAs identified in mouse embryonic stem cells (Zhao et al. 2010) (Table S6 in (Melia et al. 2016)), suggesting that many liver lincRNAs interact with chromatin-modifying enzymes beyond the ones that have been fully characterized.

**GH regulation of sex-specific liver lincRNAs.** 247 of the 4,961 liver lincRNAs showed  $a \ge 4$ -fold differential expression between male and female mouse liver at adjusted  $p \le 0.05$ ; 134 of these lincRNAs showed male-biased expression and 113 showed female-biased expression (Fig. 2-4A, Fig. 2-4B; Table S1 in (Melia et al. 2016)). Further, 59% of the sex-specific lincRNAs show high liver-specificity (Fig. 2-3B), as compared to 27% of all liver lincRNAs (4.3-fold enrichment, p = 7.04e-28; Fig. S2A in (Melia et al. 2016)), suggesting they contribute to liver-specific functions. Moreover, 195 sex-specific lincRNAs have a putative ortholog in at least one other species, including 170 with an ortholog in rat (Fig. 2-3C).

171 of the sex-specific liver lincRNAs (85 male-specific, 86 female-specific) showed a change in sex-specificity when circulating GH levels were altered, either by surgical removal of the pituitary gland (hypophysectomy) (Wauthier et al. 2010), which ablates circulating GH, or by continuous infusion of GH in male mice, which gives a female-like plasma GH profile and feminizes the expression of many sex-dependent genes in mouse liver (Holloway et al. 2006) (Fig. 2-4A, Fig. 2-4C, Table S1 in (Melia et al. 2016)). These 171 GH-dependent sex-specific lincRNAs are designated stringent sex-specific lincRNAs. An additional 17 sex-specific lincRNAs showed responsiveness to hypophysectomy when the threshold for a response was decreased from  $\geq 4$  to  $\geq 2$ -fold. However, many hypophysectomy-responsive lincRNAs are sex-independent and respond to pituitary hormone ablation in the same manner in males and females (266 out of 267 hypophysectomy-responsive, sex-independent lincRNAs; Fig. 2-4C and Table S1 in (Melia et al. 2016)), consistent with hypophysectomy affecting many non-sex-specific liver genes (Wauthier and Waxman 2008, Wauthier et al. 2010). Sex-specificity was verified for select lincRNAs by qPCR analysis of individual male and female mouse livers (Fig. 2-5A, Fig. S5 in (Melia et al. 2016)). Further characterization of the GHdependence of these lincRNAs was obtained by analysis of liver RNA from mice that were: (a) hypophysectomized, to ablate pituitary GH; (b) deficient in the GH-responsive TF STAT5, which is required for sex-specific gene expression in liver (Clodfelter et al. 2006); or (c) deficient in somatostatin, which inhibits pituitary GH release and whose knockout effectively feminizes the plasma GH profile (Adams et al. 2015) (Fig. 2-5A,

Fig. S5 in (Melia et al. 2016)). Finally, sex-specific expression was observed for two of the putative rat orthologs, whose sex- and hormone-dependent regulatory patterns in rat liver matched those seen in mouse liver, namely, female-specific expression and significant up regulation in male rat liver following continuous infusion of GH for 7 days (Fig. 2-5B).

We identified subclasses of lincRNAs based on their patterns of response to hypophysectomy (Wauthier and Waxman 2008, Wauthier et al. 2010), which either reduces or eliminates the differences in lincRNA expression between the sexes (Table 2-2). 104 (42%) of the sex-specific lincRNAs were down regulated at least 4-fold following hypophysectomy in the sex where they are more highly expressed. Thus, these genes, designated *class 1* lincRNAs, require the correspondingly matched sex-specific pituitary GH secretion profile to maintain their sex-specific expression (Table 2-2, and gene sets marked C and D in Fig. 2-4A, *right*). A second, distinct set comprised of 37 other sexspecific lincRNAs, designated *class 2*, shows significant up regulation (de-repression) following hypophysectomy in the sex where they are less highly expressed (Table 2-2). The sex-biased expression of these lincRNAs reflects their active repression by the plasma GH profile in the sex showing lower expression. Consequently, *class 2* sexspecific lincRNAs are de-repressed in the sex showing lower expression when pituitary GH secretion is ablated. Sex-specific lincRNAs: TF binding, chromatin states, and proximity to sex-specific genes. We investigated whether sex-specific lincRNA genes are enriched for sexdependent binding by TFs previously implicated in GH regulation of sex-specific proteincoding genes in mouse liver (Conforto et al. 2012, Li et al. 2012, Zhang et al. 2012, Conforto et al. 2015). Male-biased binding of STAT5, FOXA1, FOXA2 and HNF6 were significantly enriched nearby (within 10 kb) the set of 85 stringent (i.e., GH-regulated) male-specific lincRNAs, and was significantly depleted nearby the set of 86 stringent female-specific lincRNAs, when compared to a background set comprised of sexopposite and stringent sex-independent lincRNA genes (Fig. 2-6A). Similarly, the 86 stringent female-specific lincRNAs were significantly enriched, and the 85 stringent male-specific lincRNAs were significantly depleted of nearby female-biased binding by STAT5 and FOXA2. In contrast, female-specific binding of CUX2 was enriched nearby stringent male-specific lincRNAs, and male-biased binding of BCL6 was enriched nearby stringent female-specific lincRNAs. Consistent with these findings, sex-biased binding of STAT5, FOXA1/FOXA2 and HNF6 is associated with increased liver expression of nearby and correspondingly sex-biased genes (Zhang et al. 2012, Sugathan and Waxman 2013, Conforto et al. 2015), whereas the sex-biased repressors CUX2 and BCL6 preferentially repress nearby male-biased and female-biased genes, respectively (Meyer et al. 2009, Conforto et al. 2012, Zhang et al. 2012). Sex-specific binding by these factors at individual sex-specific genes is shown in Fig. 2-5C and Fig. S6 in (Melia et al. 2016). The stringent male-specific lincRNAs are significantly enriched for nearby (within 10 kb) male-biased sites of Dnase hypersensitivity (DHS) (ES = 4.78, p = 5.77e-11), and the

stringent female-specific lincRNAs are enriched for nearby female-biased DHS (ES = 5.68, p = 3.10e-9; Fig. 2-6A), consistent with these DHS (accessible chromatin regions) serving as regulatory regions for the sex-specific lincRNAs. Overall, 94 sex-specific lincRNAs have proximal sex-matched DHS; the others do not have any proximal liver DHS (45 lincRNAs), or are nearby non-sex-biased liver DHS (108 lincRNAs), suggesting they are regulated by sex-biased DHS that are distal. The control gene set, comprised of sex-independent, hypophysectomy-responsive lincRNAs (Fig. 2-4C), show no significant differential enrichments for either male-specific or female-specific TF binding or DHS, as compared to hypophysectomy-unresponsive lincRNAs that are also sex-independent. Therefore, responsiveness to hypophysectomy alone, without an associated sex-specific gene expression pattern, is not associated with proximal sex-specific TF binding and sex-specific chromatin accessibility

Next, we investigated whether sex-specific lincRNA expression is associated with specific chromatin states, which may be critical for maintaining their sex-specific expression. Specifically, we used GSEA (Subramanian et al. 2005) to determine if lincRNAs that are preferentially at a given chromatin state in one sex, but not in the other, are enriched for being either male- or female-specific lincRNAs. These analyses used liver chromatin state maps (Sugathan and Waxman 2013) based on genome-wide data for DHS and six chromatin marks for male and female mouse liver, which were used to assign defined chromatin states to each consecutive 200 bp segment of the mouse genome: the DHS and chromatin marks associated with chromatin states 1 to 3 are

indicative of inactive chromatin states, state 4 is devoid of any of the chromatin marks that were interrogated, states 5-12 represent various active and poised chromatin states, and states 13 and 14 are associated with marks for transcribed genomic regions (Fig. 2-6B). States 5 to 12 are distinguished by their patterns of DHS and activating chromatin marks: states 5, 6, 9, 10 and 11 have various combinations of classical enhancer marks (DHS, H3K4me1, H3K27ac), states 7 and 8 are marked as promoter states by the presence of H3K4me3 marks, and state 12 is a poised enhancer state (bivalent state; combination of H3K4me1 (activating) and H3K27me3 (repressive) marks).

Chromatin state enrichments were identified by GSEA for sex-specific lincRNA TSS regions and gene bodies, and then compared to those identified for matched sets of sex-specific protein-coding genes. For both male-specific and female-specific lincRNAs, we observed significant enrichment of the promoter-associated active chromatin state 7 in lincRNA TSS regions in the sex where each lincRNA set is more highly expressed, consistent with the sex-specificity of their transcription (Fig. 2-6C). A similar result was seen for sex-specific lincRNA gene bodies, and for sex-specific protein-coding genes (Fig. S7 in (Melia et al. 2016)). Further, in female liver, H3K4me1-associated enhancer states 6 and 11, and the H3K36me3-associated transcribed state 14 were enriched at female-specific lincRNA TSS (Fig. 2-6C, *right*) and gene bodies (Fig. S7 in (Melia et al. 2016)), but we did not see corresponding enrichments at male-specific lincRNAs in male liver. In male liver, H3K27me3-associated chromatin states (inactive states 1 and 2, and bivalent/poised enhancer state 12) showed significant enrichment at female-specific

lincRNA TSS regions (Fig. 2-6C, *left*). Similar patterns were seen at lincRNA gene bodies, and for protein-coding gene TSS and gene bodies (Fig. S7 in (Melia et al. 2016), *left*). However, a corresponding enrichment of H3K27me3-associated chromatin states was not found at male-specific lincRNA TSS or gene bodies in female liver, indicating H3K27me3 repressive marks are preferentially used to down regulate female-specific lincRNA genes in male liver. Finally, male-specific lincRNA TSS and gene bodies showed significant enrichment for being in chromatin state 4 in male liver. No such enrichments were seen for male-specific protein-coding genes in male liver, or for female-specific lincRNAs or protein-coding in female liver. State 4 has little or no signal intensity for any of the chromatin marks investigated (Sugathan and Waxman 2013) (similar to state 15 of (Kundaje et al. 2015)), suggesting that one or more novel chromatin marks not examined may positively regulate male-specific lincRNA expression.

Next, male-specific lincRNAs were grouped into 5 clusters using consensus clustering (Monti et al. 2003) based on the local chromatin states surrounding each lincRNA's TSS and TES in both male and female liver (Fig. S8A, S8B in (Melia et al. 2016)). Several of the clusters showed differences in chromatin state in male compared to female liver (Fig. S8C in (Melia et al. 2016), *arrows*). The most notable differences were the TES of male-specific lincRNA cluster A, which was marked by transcribed state 13 in male liver and by inactive state 2 in female liver, and the TSS of male-specific lincRNA clusters A and B, which have a higher frequency of enhancer and transcribed states and a lower

frequency of inactive states in male compared to female liver. Correspondingly, male lincRNA clusters A and B have the highest male-specificity (Fig. S8D in (Melia et al. 2016)). Further, the TSS regions of male lincRNA clusters D and E are depleted of state 11 in male compared to female liver, in line with our finding that state 11 is enriched at female-specific lincRNA TSS in female compared to male liver (Fig. 2-6C). Chromatin state differences between male and female liver were less dramatic for female-specific lincRNAs clustered in the same manner (Fig. S8E in (Melia et al. 2016)).

Finally, we investigated the proximity of sex-specific lincRNAs to sex-specific proteincoding genes, given the finding that many lincRNAs are active in regulation in *cis* (Guil and Esteller 2012). We found that 36 of the 247 sex-specific lincRNAs (15%) are proximal (within 25 kb) of a sex-specific protein-coding gene with a matched sexspecificity (Fig. S9, Table S9 in (Melia et al. 2016)), whereas only 3 sex-specific lincRNAs are proximal to sex-opposite protein-coding genes. This frequency of gene proximity is similar to that of sex-specific RefSeq protein-coding genes to other sexspecific coding genes with a matched sex-specificity (135 of 807 sex-specific coding genes, 17%). Overall, sex-specific lincRNAs are significantly enriched for proximity to a co-regulated sex-specific RefSeq protein-coding gene as compared to a sex-independent gene (ES = 4.5, p = 7.5e-11). Nevertheless, more than 200 sex-specific lincRNAs are distant from correspondingly regulated protein-coding genes, consistent with distal regulation by the latter set of lincRNAs.

## **2.E Discussion**

The gene expression landscape is shaped by the interplay between local chromatin environment and the binding of TFs and other regulators, including lincRNAs (Ernst and Kellis 2013, Sugathan and Waxman 2013). LincRNAs act by diverse mechanisms, including genome site-specific interactions with chromatin-modifying enzymes that read, write and erase histone marks (Khalil et al. 2009, Guttman et al. 2011, Engreitz et al. 2013). We characterized the mouse liver transcriptome under multiple biological conditions, which enabled us to discover 4,961 liver-expressed lincRNA genes, 59% of them novel. These lincRNA genes displayed characteristic lincRNA features, including H3K4me3 marks and RNA polymerase II binding at the promoter, H3K4me36 marks along the gene body, high tissue specificity, fewer exons than protein-coding genes, and in many cases, preferential retention in the nucleus, consistent with their proposed roles in epigenetic regulation (Khalil et al. 2009, Guttman et al. 2011). 247 liver-expressed lincRNAs showed strong (> 4-fold) sex-differences in expression, with the pituitary hormone GH serving as the major regulator of sex-biased expression, acting through both positive and negative regulatory mechanisms. Sex-specific lincRNA transcription was associated with significant enrichment for nearby, matched sex-specific open chromatin regions (sex-specific DHS regions) (Ling et al. 2010), as well as matched sex-specific binding by GH-regulated transcriptional activators and matched sex-opposite binding by GH-regulated repressors, strongly suggesting a role for these genomic regulatory regions and their bound factors in sex-specific lincRNA expression. Finally, sex-specific lincRNAs were significantly enriched for nearby protein-coding genes of the same sexspecificity, consistent with these lincRNAs contributing to regulation in *cis*; however, the majority of sex-specific lincRNAs are distant from any co-regulated protein-coding genes.

LincRNA conservation landscape. LincRNA conservation based on a multi-genome alignment of 10 supraprimate species revealed that lincRNA promoters are significantly more conserved than protein-coding promoters, while lincRNA exons are much less conserved than protein-coding gene exons. Putative orthologs of the liver lincRNAs were also identified in the same 10 species based on sequence similarity. We correctly identified orthologs for 8 of 12 experimentally validated lincRNAs characterized in multiple species, of which 6 lincRNAs are reported to interact with chromatin modifying enzymes in one or more species (Khalil et al. 2009, Zhao et al. 2010, Guttman et al. 2011). In the rat, where overall patterns of sex-specific liver gene expression are conserved (Waxman and O'Connor 2006), we identified orthologous genomic regions for 170 of 247 sex-specific mouse liver lincRNAs. We experimentally validated two predicted rat orthologs, which exhibited the same female-specificity and continuous GH-inducible expression as their mouse lincRNA counterparts.

Liver-expressed lincRNA promoters showed greater enrichment for proximal binding of liver TFs than protein-coding gene promoters, consistent with the higher conservation of liver lincRNA promoters, discussed above. Mouse liver lincRNA regulatory regions are thus more highly concentrated close to the TSS than those of protein-coding genes. This suggests that the liver lincRNA genes evolved more recently, and have not yet dispersed their regulatory regions to more distant genomic loci, as appears to have occurred with regulatory elements of liver-expressed protein-coding genes (Odom et al. 2007). Our finding of high liver lincRNA promoter conservation contrasts with other studies reporting either similar (Guttman et al. 2009) or lower (Orom et al. 2010) conservation of lincRNA promoters compared to protein-coding genes promoters in other systems. We can also anticipate that liver-expressed lincRNA and protein-coding gene promoters will be enriched for distinct sets of TF motifs, as was found for human lincRNAs (Alam et al. 2014). Finally, we found that multi-tissue expressed liver lincRNAs show an even higher level of proximal promoter conservation than liver-specific lincRNAs. Tissue-specificity could therefore be a confounding variable when comparing promoter conservation between lincRNA and protein-coding genes.

**Regulation of sex-specific lincRNAs by GH.** 171 of the 247 sex-specific lincRNAs were responsive to a change in plasma GH profiles by either pituitary ablation (hypophysectomy) or continuous infusion of GH in male mice, with the latter leading to down regulation of many male-specific liver lincRNAs and up regulation of many female-specific lincRNAs, as also occurs with many sex-specific protein-coding genes (Holloway et al. 2006, Wauthier et al. 2010). Two major classes of sex-specific lincRNAs were identified based on responses to hypophysectomy (Table 2-2). Class 1 sex-specific lincRNAs are positively regulated by the sex-specific GH secretion pattern of the sex where they are more highly expressed, and consequently, are down regulated by pituitary hormone ablation. In contrast, class 2 sex-specific lincRNAs are repressed by the GH secretion pattern of the sex where they are less highly expressed, and consequently, are up regulated by pituitary hormone ablation. These findings suggest that the class 1 lincRNAs are regulated by GH-dependent TFs that confer positive regulation, e.g. STAT5, while the class 2 lincRNAs are regulated by sex-specific, GH-dependent repressors, such as BCL6 and CUX2 (Conforto et al. 2012, Zhang et al. 2012). Supporting this, we found that, overall, male-specific lincRNAs are enriched for nearby male-biased binding of the pioneer factors FOXA1 and FOXA2 (Friedman and Kaestner 2006, Zaret and Carroll 2011), which can induce chromatin opening (Sugathan and Waxman 2013), and for male-biased binding of the GH-regulated transcriptional activators STAT5 (Zhang et al. 2012) and HNF6 (Conforto et al. 2015), which may cooperate with FOXA1 and FOXA2 to induce male-specific lincRNA expression. Malespecific lincRNAs also showed enrichment for female-specific binding of CUX2 (Conforto et al. 2012), which represses male-biased gene transcription in female liver. Female-specific lincRNAs were significantly depleted of nearby male-biased transcriptional activator binding events, but were enriched for nearby male-biased transcriptional repressor (i.e., BCL6) binding events, which serves as a mechanism to suppress female-biased lincRNAs in male liver; and were significantly enriched for female-biased binding by STAT5 and FOXA2, which are proposed to activate transcription of the associated female-specific lincRNAs. Of note, only 7 of the 171 GHregulated sex-specific lincRNAs were expressed in other tissues with high STAT5 activity, i.e. T cells and mammary gland (Furth et al. 2011, Mahmud et al. 2013),

indicating that STAT5 activity alone is not sufficient to drive expression of these genes (data not shown). Two of the 7 co-expressed lincRNAs have a nearby (within 10 kb) STAT5 binding site in liver (c.f. nearby STAT5 binding for 107 of the 171 GH-regulated, sex-specific liver lincRNAs): LiverLincs\_chr14\_3799 (expressed in mammary gland, Tregs and Th2 cells, but not Th1 cells), and LiverLincs\_chr6\_1886 (expressed in all three T cell subtypes).

Chromatin state analysis revealed striking differences between male-specific and femalespecific lincRNAs regarding sex differential enrichment for certain active and repressive chromatin states. Whereas the promoter-associated active chromatin state 7 was enriched, both at male-specific lincRNAs in male liver and at female-specific lincRNAs in female liver, enhancer-associated states 6 and 11 showed significant enrichment only for femalespecific lincRNAs in female liver. State 11, which is marked by H3K4me1 but not by H3K27ac or DHS (Fig. 2-6B), may be viewed as an inactive enhancer state (Creyghton et al. 2010, Rada-Iglesias et al. 2011). However, as state 6 is characterized by three active marks, H3K27ac, H3K4me1, and DHS, and is thus an active enhancer state, its lack of enrichment at male-specific lincRNAs suggests that male-specific lincRNA expression primarily involves distal male-specific enhancers. We also found that female-specific lincRNA TSS regions were enriched in male compared to female liver for three H3K27me3-associated states, namely, inactive states 1 and 2, and poised enhancer state 12. A corresponding enrichment of repressive chromatin states was not found at malespecific lincRNAs in female liver, indicating preferential usage of H3K27me3 as a

mechanism to repress female-specific lincRNAs in male liver. This finding is reminiscent of the prevalence of H3K27me3 marks at a subset of the most highly female-specific genes in male liver (Sugathan and Waxman 2013), and suggests that GH-regulated malespecific lincRNAs that interact with PRC2, which catalyzes formation of H3K27me3 marks, may repress female-specific lincRNAs and certain female-specific protein-coding genes by a common mechanism.

Relationship between sex-specific lincRNAs, sex-specific chromatin states and protein-coding gene expression. The enrichment of sex-specific lincRNAs for nearby and correspondingly sex-biased binding by GH-dependent transcriptional activators and sex-opposite binding by GH-dependent repressors suggest a role for these TFs in the proximal regulation of sex-specific lincRNA transcription, in addition to their crucial roles in regulating sex-specific protein-coding gene expression (Ling et al. 2010, Conforto et al. 2012, Zhang et al. 2012, Sugathan and Waxman 2013). One such TF, STAT5, is a key GH-responsive factor whose deletion in male mouse liver leads to widespread disruption of sex-specific gene expression (Clodfelter et al. 2006). Sexspecific lincRNAs activated by STAT5 (c.f., Fig. 2-5A, Fig. S5, Fig. S6 in (Melia et al. 2016)) and other GH-dependent TFs may, in turn, act as scaffolds for chromatin regulators to establish or maintain sex-specific chromatin states. Such sex-specific lincRNAs could thus link two layers of gene regulation, GH-dependent TFs and GH regulation of liver chromatin states. The proposal that sex-specific lincRNAs serve as mediators between the effects of sex-specific plasma GH profiles, and the GH-regulated TFs that they regulate, in establishing the distinct chromatin states found in male and female liver (Sugathan and Waxman 2013) is supported by the unexpectedly long time required for some sex-specific liver genes to respond to feminization of circulating GH profiles by continuous growth hormone treatment (response times ranging up to a week or longer) (Holloway et al. 2006). Moreover, 350 of the 4,961 liver-expressed lincRNAs characterized here are reported in mouse embryonic stem cells to bind PRC2, which catalyzes the deposition the repressive histone mark H3K27me3 (Zhao et al. 2010), further supporting the role of lincRNAs in regulating chromatin states via chromatin-modifying enzymes (Guttman et al. 2011). Further work will be required to establish the functional roles of the sex-specific liver lincRNAs described here and any contributions they may make to the regulation of sex-specific liver gene expression.

In conclusion, by using livers collected under diverse biological conditions, we characterized gene structures, expression patterns, and species conservation for ~5,000 liver-expressed lincRNAs, including ~3,000 novel lincRNA genes. Collectively, these liver lincRNAs show unexpectedly high promoter conservation and elevated TF binding activity in the proximal promoter region compared to protein-coding genes. We focused on a population of condition-specific lincRNAs showing significant sex-specific, GH-regulated expression. By integrating genome-wide chromatin state maps and genome-wide binding site data for key GH-regulated TFs, we showed that sex-specific liver lincRNA expression is tightly correlated with the surrounding chromatin environment and TF binding patterns. LincRNA orthologs identified across species include at least

two rat liver orthologs with matched sex-specificity and responsiveness to sex-dependent plasma GH profiles. These findings illustrate how multi-omics data can be integrated to elucidate molecular mechanisms underlying condition-specific lincRNA regulation.

## 2.F Acknowledgements

I would like to thank the following laboratory members who contributed tissue samples and/or RNA-seq samples: Nicholas J. Lodato, Dana Lau-Corona, Alexander Suvorov, Jeannette Connerney, and Tara L. Conforto. I also thank Gracia Bonilla, Andy Rampersaud and George F. Steinhardt IV of this laboratory for ChIP-seq peak discovery for FOXA1 and FOXA2 (GB), CEBPA and CEBPB (AR), and HNF6 (GFS).

Genes and transcripts	lincRNA count	average # exons	exon length (nt)	intron length (nt)	mature transcript length (nt)
All liver-expressed lincRNAs	4,961	1.6	1,075	5,255	1,866
Multi-exonic	1,179	3.4	998**	5,255	3,393*
Mono-exonic	3,782	1.0	1,381	-	1,390
Sex-specific	247	2.5	1,226	5,436	3,792
RefSeq coding	19,951	10.1	294	5,158	2,897
RefSeq noncoding	3,156	3.7	384	6,169	1,472

Table 2-1. Exon and introns of novel lincRNA, RefSeq protein-coding, and RefSeq noncoding genes

Comparisons for multi-exonic lincRNAs: \*\*, p < 1e-15 vs. RefSeq coding RNAs, and

\*, p = 1.2e-6 vs. RefSeq coding RNAs (Student's t-test).

Table 2-2. Sex-specific mouse liver lincRNAs that respond to hypophysectomy (hypox)

Sex specific	Response to male hypox	Response to female hypox	Hypox, $ fold change  \ge 2$	Hypox, $ fold change  \ge 4$
Incrina gene class			Adjusted p-value $\leq 0.05$	
		Sex-specific lincRNAs (number)		
Male class 1	Down	-	65	61
Male class 1A	Down	-	46	45
Male class 1B	Down	Down	17	15
Male class 1C	Down	Up	2	1
Male class 2	-	Up	12	10
Male class 2A	-	Up	9	9
Male class 2B	Up	Up	3	1
Female class 1	-	Down	44	43
Female class 1A	-	Down	33	33
Female class 1B	Down	Down	11	10
Female class 1C	Up	Down	0	0
Female class 2	Up	-	27	27
Female class 2A	Up	-	18	20
Female class 2B	Up	Up	9	7
## Figure 2-1. Liver-expressed lincRNAs.

Liver-expressed lincRNAs. (A) Scheme for filtering 7,088 intergenic transcripts to remove transcripts with very low expression, small RNAs, and transcripts that resemble protein-coding genes. The set of 4,961 liver-expressed lincRNAs was analyzed to flag identify small RNAs and pseudogenes, and sequences with ORFs > 150 amino acids, giving a set of 4,454 stringent lincRNAs. (B) Protein-coding potentials across 29 mammals, based on genes that are represented in the multiple-genome alignment of the 29 mammals. Data shown are based on 3,036 of 4,554 liver-expressed lincRNAs without RefSeq annotations (i.e., non-RefSeq lincRNAs, a subset of the 4,961 liver-expressed lincRNAs), 2,217 RefSeq ncRNAs, and 6,567 RefSeq coding RNAs. An increasing positive score along the X-axis indicates greater similarity of the codon substitution pattern to known protein-coding genes. The protein-coding potential filter applied to liver-expressed non-RefSeq lincRNAs was more stringent than the one used to identify RefSeq ncRNAs. (C) Distribution of RNA expression levels in total vs. nuclear RNA fractions for each of the indicated gene sets (19,951 RefSeq protein-coding genes, 1,557 RefSeq ncRNA genes expressed in at least one liver sample, and 4,554 non-RefSeq lincRNAs). Data shown are based on the maximum expression level of each gene across RNA-seq datasets whose biological conditions are represented in both total liver RNA (n=20 data sets) and nuclear liver RNA (n=20 data sets). The 4.554 lincRNAs show a greater difference in expression in nuclear compared to total liver RNA fractions than the other two gene sets. P-values are based on student's t-test. (D) Principal component analysis of lincRNA expression per sample based on read counts over exons, based on analysis of the 500 most-varying lincRNAs across samples. Each data point represents a single RNA-seq dataset. The largest variance across datasets (first principal component,  $\sim$ 70% of total variance) separates nuclear from total RNA samples, and the fifth principal component (~3% of total variance) separates male from female samples. (E) Tissuespecific expression for the set of 2,171 lincRNA genes expressed in liver at  $\geq$  1 FPKM (also see Fig. S3 in (Melia et al. 2016)). Shown is the number of tissues, other than liver, where each lincRNA is expressed at  $\geq$  1 FPKM. 1,437 of the 2,171 lincRNAs are expressed at  $\geq$  1 FPKM only in liver, and 30 lincRNAs are expressed at  $\geq$  1 FPKM in all 22 tissues, of which 23 lincRNAs are known RefSeq noncoding RNAs (Table S2 in (Melia et al. 2016)).





## Figure 2-2. Conservation of lincRNA promoters.

(A) Per-base conservation analysis of different gene-centric regions of lincRNAs, RefSeq coding genes, random background sequences, and control promoter regions (1 kb downstream of TSS; see text). The thickness of the background curve represents the 95% confidence interval based on 100 random samplings of gene structure-matched and length-matched untranscribed intergenic regions, which mimic our lincRNA gene set. Curves for protein-coding introns and lincRNA introns overlap. LincRNA promoters showed significantly higher conservation than protein-coding promoters (p < 1e-15, Kolmogorov-Smirnov). A background set of promoter-associated sequences, using 1 kb downstream of the lincRNA TSSs, was used as a benchmark to compare promoters of lincRNA genes to those of RefSeq protein-coding genes. (B) Per-base conservation analysis of the promoter regions of the most highly expressed (top 33%) liver-specific genes (222 lincRNAs, 282 protein-coding genes; Fig. S4A, S4B in (Melia et al. 2016)) (top) and multi-tissue expressed genes (289 lincRNAs, 3,619 protein-coding genes; Fig. S4A, S4B in (Melia et al. 2016)) (bottom). In both panels, lincRNA promoters showed consistent high conservation compared to protein-coding gene promoters (dashed vs. solid blue lines); and multi-tissue promoters, for both lincRNAs and protein-coding genes, showed greater conservation than the corresponding liver-specific promoters. (C) Aggregate plots of mean per-base conservation scores surrounding lincRNA and RefSeq protein-coding gene TSS (*left*) and TES (*right*). See Fig. S4C-S4E in (Melia et al. 2016) for corresponding plots of select, individual lincRNAs. (D) TF occupancy frequency 1 kb (left) and 25 kb (right) upstream of sets comprised of 114 FPKM-matched liver-specific lincRNA and liver-specific protein-coding genes, whose expression level is >0.52 FPKM (see Fig. S4A in (Melia et al. 2016)). The TF occupancy rate at lincRNAs promoters was significantly higher (\*: p < 0.05, Binomial exact test) than at protein-coding gene promoters for 11 out of 12 TFs examined, but not for DHS.





#### Figure 2-3. Liver lincRNA species conservation and tissue specificity.

(A) Heat map showing species conservation of 3,479 species-conserved liver lincRNAs, i.e., lincRNAs that have an ortholog in at least one of the 10 species investigated. Orthologs were required to span > 10% of the full length, mature mouse lincRNA transcript and be > 600 bp in length (see Methods). (B) Heat map displaying tissue specificity of 247 sex-specific lincRNAs. The map is based on the summation of the expression levels of each lincRNA across the 22 indicated mouse tissues, with the color intensity indicating the fraction of expression seen in each tissue. Data shown are based on two RNA-seq samples for each tissue. Liver data are also presented as two samples, one representing total RNA samples and the nuclear RNA samples (see Methods). 146 (59%) of the sex-specific lincRNAs have at least 60% of their cumulative expression across the tissue panel concentrated in liver (red); 105 (42%) have 80% of their cumulative expression concentrated in liver. (C) Subset of the heat map shown in (A), where species conservation is displayed for the set of 195 sex-specific liver lincRNAs conserved in one or more species. Magnitude of the sex-difference in lincRNA expression is as indicated (FC, fold-change, corresponding to the ratio of male and female liver expression levels of each lincRNA). In data not shown, 105 (76%) of the 138 multi-exonic sex-specific lincRNAs have an ortholog in at least one of the 10 species.

Fig. 2-3



61

## Figure 2-4. GH regulation of sex-specific liver lincRNAs.

GH regulation of sex-specific liver lincRNAs. (A) Heat map of 247 sex-specific liver lincRNAs. The first six heat map columns present the sex-specificity of the lincRNAs in four total liver RNA datasets (T1-T4) and in two nuclear RNA datasets (N1, N2), the next three columns show the changes in liver lincRNA expression after continuous GH treatment of male mice for 1, 4 or 7 days, and the last two columns show the changes in expression following hypophysectomy (Hypox, Hx) of male (M) and female (F) mice. Color bar specifies the log<sub>2</sub> gene expression ratios for the indicated sets of liver RNA samples. LincRNA genes marked A (right) are female-biased genes that are up regulated following continuous GH treatment, and genes marked B are male-biased genes that are down regulated following continuous GH treatment (columns D1, D4, D7). Genes marked C are male-biased genes that require the male plasma GH pattern to maintain sexspecificity (and hence are down regulated in M-Hx livers), and genes marked D are female-biased genes that require the female plasma GH pattern to maintain their sexspecificity (and hence are down regulated in F-Hx livers). (B) MA plot presenting sex ratio vs. gene expression level, in FPKM, for 247 sex-specific lincRNAs and for 240 sexspecific protein-coding genes showing > 4-fold sex-differences. Values are based on the maximum FPKM value (x-axis) and the maximum fold-difference between male and female samples (y-axis) across the datasets considered in this study. Male-specific lincRNAs show a similar range of sex ratios but lower FPKMs than male-specific protein-coding genes; whereas, there are many fewer highly female-specific lincRNAs than highly female-specific protein-coding genes. (C) Venn diagrams showing overlaps between 247 sex-specific lincRNA genes, 411 hypophysectomy-responsive lincRNA genes, and 252 continuous GH treatment-responsive lincRNA genes. The thresholds for sex-specificity and responsiveness to hypophysectomy were set at  $\geq$  4-fold and adjusted p-value  $\leq 0.05$  (except as noted in Methods); the threshold for responsiveness to continuous GH treatment was set at  $\geq$  2-fold and adjusted p-value  $\leq$  0.05. 171 of the 247 sex-specific genes are responsive to either continuous GH or hypophysectomy (78 + 66 + 27). Of note, the size of the non-sex-specific gene sets that respond to continuous GH treatment or hypophysectomy are sensitive to the choice of significance level cutoff. Thus, of the 147 sex-independent lincRNAs that respond to continuous GH treatment (109 + 38), 53 (36%) are sex-specific when the threshold for sex-specificity is relaxed, from  $\geq$  4-fold to  $\geq$  2-fold. Robustness of sex specificity: 75% (129), 57% (97) or 38% (65) of the 171 sex-specific, GH-responsive lincRNAs exhibit consistent sex-specificity in at least 2, 3 or 4 datasets at fold change  $\geq 2$  and adjusted p  $\leq 0.05$ . The finding of 267 (=38 + 229) hypophysectomy-responsive lincRNAs that are sex-independent presumably reflects regulation by pituitary-dependent hormones other than GH, a common feature of non-sex-specific protein-coding genes (Wauthier et al. 2010). These 267 genes were used as a control gene set in Fig. 2-6A.



### Figure 2-5. Regulation of select sex-specific liver lincRNAs.

Regulation of select sex-specific liver lincRNAs. (A) qPCR analysis of individual malespecific (top) and female- specific (bottom) lincRNAs. Sex-specific expression of the male lincRNA is lost when GH pulsatile secretion is abolished, as seen in hypophysectomized mice (where GH is ablated) and in somatostatin-knockout mice (where the plasma GH pattern is female-like). This lincRNA is also regulated by the GHresponsive STAT5, as its expression becomes sex-independent in hepatocyte-specific STAT5-knockout mice. The female-specific lincRNA shows dependency on the continuous secretion of GH: its expression is ablated in hypophysectomized female mice and is highly induced in somatostatin-knockout male mice; this lincRNA is not dependent on STAT5 for its expression. (B) qPCR analysis of rat liver RNA mapping to rat genomic regions (rn4) orthologous to two sex-specific mouse lincRNAs. Data shown are for RNA samples isolated from individual rat livers (mean + SD for triplicate reactions), and indicate that the rat lincRNA orthologs are expressed in a female-biased manner and are inducible to female-like levels in livers of male rats given a continuous infusion of GH for 7 days (male + GH). Rat qPCR primers are shown in Table S10 in (Melia et al. 2016). (C) Genome browser screenshots for one male-specific (*left*) and one female-specific (right) lincRNA. The gene models track (orange) shows the gene structures for each lincRNA, RNA-Seq tracks show the expression level in male (blue) and female (pink) liver, chromatin state map tracks show the chromatin state separately in male and female liver (see Fig. S6 in (Melia et al. 2016) for color legend), DHS and TF binding tracks show open chromatin regions and binding sites that are either male-enriched (blue rectangle), female-enriched (pink rectangle) or sex-independent (grav rectangle), and chromatin mark tracks show signal intensity for either H3K4me3, H3K27ac and H3K4me1 in male (blue) and female (pink) liver. For each sex-specific lincRNA, we observed: matched sex-biased binding for activating TFs (STAT5, FOXA1, FOXA2, HNF6); sex-opposite binding by the repressors BCL6 and CUX2; and matched sexdifferential active chromatin mark signals. Only multi-exonic gene models are shown for LiverLincs chr6 1718 (right).





### Figure 2-6. Sex-specific TF binding and chromatin states at sex-specific lincRNAs.

(A) Enrichment and depletion scores of stringent sex-specific lincRNAs as well as a control gene set (comprised of 38+229 = 267 hypophysectomy-responsive, sexindependent lincRNAs; Fig. 2-4C) for nearby (10 kb) sex-biased DHS and six GHresponsive TFs implicated in sex-specific liver gene expression. \*, p < 5e-2; \*\*, p < 5e-4; and \*\*\*, p < 5e-6. The background used for the enrichment was the set of sex-opposite + stringent sex-independent lincRNAs for stringent sex-specific lincRNAs or hypophysectomy-unresponsive, sex-independent lincRNAs for the control gene set. (B) Emission probabilities across six chromatin marks and DHS for each of 14 chromatin states, as reported previously (Sugathan and Waxman 2013). The darker the blue color, the greater the emission frequency of the indicated chromatin mark or DHS in a given chromatin state. (C) GSEA analysis of chromatin state enriched in either male-specific or female-specific lincRNA TSS regions. 4,961 liver lincRNAs are ordered along X-axis by male/female expression ratio. Input gene lists are displayed below the X-axis, with each vertical mark indicating a lincRNA whose gene body is in the indicated chromatin state in male but not female liver (*left*), or in female but not male liver (*right*). As an example, results at left show that, in male liver, states 4 and 7 are enriched at male-specific lincRNA TSS (positive NES values), and states 1, 2 and 12 are enriched at femalesspecific lincRNA TSS (negative NES values). See Fig. S7 in (Melia et al. 2016) for corresponding GSEA analysis of lincRNA gene bodies and protein-coding gene TSS and gene bodies.



# Chapter 3 Genetic factors increase the variability of sex-specific gene patterns in Diversity Outbred mouse liver

# **3.A Abstract**

The mammalian liver transcriptome features widespread sex differences affecting  $\sim 1.000$ genes, many of which are implicated in sex differences in drug and lipid metabolism, as well as disease susceptibility. Growth hormone (GH) is a major regulator of sex-biased gene expression through its sex-differential activation of various transcriptional regulators. Genetic polymorphisms have been shown to impact the transcriptional regulation of select sex-biased genes. Global assessment of the impact of genetic factors on sex-biased gene expression, however, is lacking, Here, we show that genetic factors, working in tandem with GH, have widespread effects on individual variability in sexbiased gene expression in livers of Diversity Outbred (DO) mice, an outbred population derived from 8 founder strains. Significant associations between single nucleotide polymorphisms (SNPs) and gene expression (eQTLs) were identified, a subset of which correlated with sex-biased genes and showed stronger association in one sex, as compared to the other. Sex-biased gene expression was highly variable between DO founder strains. Remarkably, 46% of the cases of gain or loss of sex-specific gene expression in individual founder strains can be explained by genetic modifiers identified by eQTLs. eQTLs with strong sex-biased associations are enriched for genetic modifiers identified for highly sex-specific genes with matching sex-specificity. eQTL regions for sex-specific genes are enriched for correspondingly sex-biased open chromatin regions

(DHS) and binding sites for the transcriptional activator STAT5. Sex-biased binding of the transcriptional repressor BCL6 was enriched at both *cis* and *trans* eQTL regions for female-specific genes in male liver. A subset of co-regulated gene clusters identified based on overlapping eQTLs encompassed highly correlated genes from different chromosomes. Thus, genetic factors are pervasive in regulating sex-biased genes in liver, which has important consequences to discern individual variability of liver phenotype and disease outcome.

## **3.B Introduction**

Sex differences in mammalian gene expression are not limited to reproductive tissues, but also occur somatic tissues (Rinn and Snyder 2005), notably the liver, as seen in mouse (Clodfelter et al. 2007, Renaud et al. 2011, Conforto and Waxman 2012), rat (Wauthier and Waxman 2008) and human (Tanaka 1999, Zhang et al. 2011). ~1,000 genes are differentially expressed between male and female mouse liver (Conforto and Waxman 2012), imparting sex differences in various liver functions, such as metabolism of drugs (Anthony and Berg 2002, Guengerich 2006, Zanger and Schwab 2013), other xenobiotics (Klaassen and Aleksunes 2010, Liu et al. 2013), and fatty acids (Wang et al. 2011), as well as disease susceptibility (Lu et al. 2009, Tessari et al. 2009, Baik et al. 2011, Zhang et al. 2011). Sex differences in the liver transcriptome are widespread and characterize both protein-coding transcripts (Conforto and Waxman 2012) and long noncoding RNAs (IncRNA) (Melia et al. 2016). Growth hormone (GH), secreted by the pituitary gland, has long been known to be a key regulator of sex-biased gene expression (Jansson et al. 1985,

Jaffe et al. 2002, Waxman and O'Connor 2006) via its sex-dependent pattern of pituitary secretion: pulsatile in males and nearly continuous in females (Veldhuis et al. 2001, Veldhuis and Bowers 2003). The resultant sex differences in plasma GH profiles lead to the sex-differential activation of the JAK2/STAT5 signaling pathway. GH-activated STAT5, in turn, works with other GH-dependent TFs to propagate sex-differences in liver transcription through sex-biased binding to liver chromatin (Laz et al. 2007, Conforto et al. 2012, Zhang et al. 2012, Conforto et al. 2015). Sex differences are also evident in the liver epigenome, which includes sex differences in open chromatin regions (DNase I hypersensitive site; DHS) (Ling et al. 2010), chromatin marks (Sugathan and Waxman 2013), and DNA methylation (Reizel et al. 2015), all of which are tightly linked to the transcription of sex-specific genes.

Genetic modifiers can contribute to the regulation of sex-biased genes (Krebs et al. 2003, Krebs et al. 2005, Krebs et al. 2009, Krebs et al. 2012). In one well studied example, male-biased expression of *Cyp2d9* is abolished by a SNP in a 5' flanking regulatory region (Sueyoshi et al. 1995). In a second example, Regulator of Sex-Limitation (*Rsl*) gene was shown to repress several sex-specific genes, and was associated with activation and repression of other sex-specific genes *via* indirect mechanisms (Krebs et al. 2009, Krebs et al. 2012). Genetic polymorphisms contribute to individual differences in several drug-metabolizing enzymes (Hines et al. 2008), many of which are encoded by sexspecific genes (Mugford and Kedderis 1998, Waxman and Holloway 2009). Genomewide studies to discover significant associations between genetic variants and liver gene expression have identified many expression quantitative trait loci (eQTLs) for liverexpressed genes (Gatti et al. 2007, Gatti et al. 2010, Chick et al. 2016, Gatti et al. 2017, Tyler et al. 2017). None of these studies, however, assessed the impact of eQTLs on the individual variability of sex-biased gene expression. A global assessment of the extent to which genetic factors alter sex-biased gene expression is still lacking. Understanding the source of variability in sex-biased gene expression in liver has far-reaching implications, as sex-biased genes contribute to liver diseases with known sex-differential susceptibility, including hepatocellular carcinoma (Ruggieri et al. 2010, Baik et al. 2011), polygenic dyslipidemia (Bittner 2006) and coronary artery disease (Zhang et al. 2011).

To ascertain the global impact of genetic regulation on sex-specific gene expression, we utilized the Diversity Outbred (DO) mouse model (Churchill et al. 2012, Schmidt 2015), an outbred population derived from eight inbred mouse strains. DO mice have high natural allelic variance (Svenson et al. 2012, Chesler et al. 2016) and diverse phenotypes (Bogue et al. 2015, Odet et al. 2015), which may lead to finer genetic mappings for liver-specific traits. Further, genome sequences from all DO founder strains are available (Keane et al. 2011), allowing genotypes in the DO mouse to be traced back to select founder strains. Using the DO mouse model, we find unexpectedly high rates of gain and loss of sex-specific gene expression across DO founder strains. ~13,000 autosomal eQTLs – many novel and including eQTLs for long noncoding RNAs (lncRNAs) – were identified in DO mouse liver. A subset of these eQTLs show sex-bias genetic associations, which we used to explain the gain or loss of sex-specific gene expression in

DO founder mice. Sex-specific binding of GH-regulated TFs and sex-specific open chromatin regions (DHS) were shown to be enriched at eQTL regions identified for genes with the expected sex-specificity. Finally, co-regulated genes were identified on the basis of overlapping eQTLs, a subset of which contain highly correlated genes from different chromosomes. These findings establish the pervasive role of genetic regulation in increasing the variability of sex-specific gene expression across DO founder strains. Understanding the source and extent of this variability has important implications for our understanding of individual variability as it relates to sex-differences in liver disease and function.

# **3.C Materials and methods**

**DO mouse liver samples and analyses.** Liver RNA sequencing expression data and genotyping data based on SNP arrays was downloaded from GEO for 438 individual DO mice using accession numbers GSE45684 (Munger et al. 2014, Gu et al. 2016) and GSE72759 (Chick et al. 2016). These datasets are derived from 219 male and 219 female DO mice, which were further divided into tissues from mice fed a high fat diet (107 males and 109 females) or a standard chow diet (112 males and 110 females). 264 DO samples were genotyped at 7,854 SNPs using the Mouse Universal Genotyping Array (MUGA) (Consortium 2012) and 174 were genotyped using a higher density array, MegaMUGA (Morgan et al. 2015), encompassing 77,725 SNP probes. **Haplotype reconstruction.** I used quantile normalization, implemented in DOQTL (Gatti et al. 2014), to normalize microarray intensity values across batches with default parameters. I

further used DOQTL to calculate from the normalized microarray intensity values at each SNP probe the founder strain that the SNP is most likely inherited from, i.e. the founder haplotype. Two founder haplotypes, referred to as founder diplotype, are possible at each SNP locus. Thus, there are 36 possible founder diplotypes for a population generated from 8 founder strains: 8 homozygous diplotypes and 28 heterozygous diplotypes. Briefly, DOQTL uses the intensity of each SNP probe to generate a probabilistic estimate for each diplotype state at each SNP in each DO mouse using a hidden Markov model (HMM), where the hidden states are the diplotype states. DOQTL assigned each SNP locus to the diplotype state with the highest posterior probability. **Individual mouse** genome reconstruction. I used Segnature (Munger et al. 2014) to construct an individual diploid genome for each DO mouse using founder haplotypes that were inferred at the previous step. Recombination boundaries are simply the midpoint between neighboring SNPs that were assigned to different founder diplotypes. For each recombination block, Sequence recapitulates the founder genome of interest by incorporating high quality SNPs and small indels (<100 bases) that were found in that particular genome, downloaded as a vcf file from the Sanger Mouse Genome project (release 1211, (Keane et al. 2011)). RNA-Seq read mapping. I mapped RNA-Seq reads from each individual DO mouse liver sample to its diploid genome, containing both paternal and maternal allele sequences, using Tophat2 (Langmead et al. 2009, Kim et al. 2013) with default parameters (Fig. S3-1).

Other mouse liver RNA-Seq samples and data analyses. I downloaded three publicly available liver gene expression datasets for the eight DO mouse founder strains: 1) a microarray dataset comprised of 96 samples, six replicates/sex/strain, assayed on the Illumina Sentrix Mouse-6 V1.1 platform (http://cgd.jax.org/gem/strainsurvey26/v1); 2) an RNA-Seq dataset with 128 male liver samples, six replicates/strain/diet, where the two options for diet are standard chow and high fat diet (GEO accession GSE45684) (Munger et al. 2014, Gu et al. 2016); and 3) an RNA-Seq dataset with 12 male and 12 female C57Bl/6J mouse livers (GEO accession GSE59222) (Lowe et al. 2015). All of the RNA-Seq datasets were mapped to the genome that matched their respective strain generated by Seqnature, based on SNPs/indels retrieved from the Sanger Mouse Genome project (release 1211) (Keane et al. 2011), using Tophat2 with default parameters. A published mouse liver RNA-Seq dataset from the CD-1 mouse strain, consisting of three pools of male and three pools female liver samples, was downloaded from GEO accession GSE98586 (Lau-Corona et al. 2017). These samples were mapped to the reference genome using Tophat2 with default parameters (Fig. S3-1).

**Expression quantification.** Gene expression was quantified by counting sequence reads that overlap any exon by at least one bp using featureCounts (Liao et al. 2014). For RNA-Seq datasets that were mapped to a haploid genome, e.g. the C57Bl/6J reference mouse genome and any of the eight DO mouse founder strains, only unique reads were used, whereas for datasets that were mapped to a diploid genome, e.g. individual DO mouse genomes, the best mapped location for each read was used. The restriction of only using

unique reads for diploid genomes would limit the read counts to only include those reads that mapped to genomic locations where the paternal and maternal allele differ, and thus could substantially underestimate expression of any given gene. Further, the gene expression level obtained for diploid genomes is based on the total number of reads that overlap the counted regions in either the paternal or maternal allele. Reads counts were then transformed to fragments per kilobase of exon per million reads mapped (FPKM) for downstream analysis. For FPKM calculation on diploid genomes, exon lengths are based on the average length of exons from the two alleles of each gene.

**Sex-specificity of gene expression.** I used three liver expression datasets, described above, to establish the sex-specificity of protein-coding genes: 1) RNA-Seq dataset of male and female CD1 mouse livers, 2) RNA-Seq dataset of male and female CD1 mouse livers, 2) RNA-Seq dataset of male and female CD1 mouse livers, 2) RNA-Seq dataset of male and female CD1 mouse livers from the eight DO mouse founder strains. I used edgeR (Robinson et al. 2010) to perform differential analysis for the two RNA-Seq datasets; and limma (Smyth 2004) was used for the microarray dataset with default parameters. For genes with multiple microarray probes, we chose the probe with the smallest adjusted p-value. A total of 1,033 protein-coding genes showed a male/female gene expression |fold-change| > 2 at FDR < 0.05 for any RNA-Seq dataset or |fold-change| > 1.5 at FDR < 0.05 for the microarray dataset, and were designated sex-specific (Table S3-1). 168 multi-exonic, intergenic liver-expressed lncRNAs were designated sex-specific, as described in Chapter 4.

Gene expression variability across DO founder strains. Inter-strain gene expression variability across the 8 DO mouse founder strains was quantified for each gene using RNA-seq data for 8 male livers per strain, as follows: [standard deviation of (mean gene expression value for each of the 8 strains)] / [average of (the mean gene expression value for each of the 8 strains)]. Intra-strain gene expression variability was quantified for each gene in each strain as follows: [standard deviation of (gene expression for n = 8 male livers in strain *s*)] / [mean of (gene expression value for n = 8 male livers in strain *s*)]).

**eQTL mapping.** eQTL analysis was carried using liver gene expression data from 438 DO mice, comprising 219 male and 219 female mice that were either fed a high fat or standard chow diet (Fig. S3-1). We used the additive haplotype model in DOQTL (Gatti et al. 2014) to perform QTL mapping by regressing each gene's expression level on the estimates of each of the eight founder strain's contribution, i.e., founder diplotype probability, at each SNP marker. Adjustments to account for relatedness amongst DO mice, batch, sex, diet, and interaction between sex and diet were included in the regression as additive covariates, written as the following in R: model.matrix(~sex + diet + diet\*sex + generation + batch, data = data). The strength of the association between gene expression and genotype is given as a likelihood ratio (LOD), which is the -log10(p-value) when comparing the full model to the null model, where the latter model excludes diplotype probabilities. QTL mapping also gives eight regression coefficients, whose magnitudes reflect the effect of the founder alleles at each SNP marker. A positive

regression coefficient indicates the genetic variant is associated with high expression of the gene of interest, whereas a negative regression coefficient indicates the genetic variant is associated with low expression of the gene of interest. The DO founder strain with the largest gene expression alteration by genetic factors identified in each eQTL, as indicated by the largest absolute value of the regression coefficients, is designated as 'the regulating strain'. A Bayesian credible interval, defined as 95% of the region under the LOD^10 curve, as implemented in DOQTL (Gatti et al. 2014), was defined for each eQTL; this interval delineates the genomic location where the highest association occurs, as defined by the uppermost 5% area under the peak of the LOD score curve.

**Significance level.** A genome-wide p-value for each association was determined by doing 1,000 permutations on the gene expression data. To assemble a list of associations between gene expression and SNPs, I recorded the SNP with the highest LOD score for each gene; subsequent high-scoring SNPs were only retained if their p-value < 0.05 and if they are on a different chromosome than the SNP with the highest LOD score. An FDR correction was then applied to the genome-wide p-values. Any eQTL with FDR < 0.05 was deemed significant. To identify genetic variants whose association with gene expression was only found in one sex, we repeated the eQTL mapping analysis twice more, once using only male DO liver samples, and a second time using only female DO liver samples. Gene expression pre-processing. Gene expression levels in FPKM were transformed into normal scores using the inverse normal transformation in DOQTL (Gatti et al. 2014) before they were used for eQTL mapping. eQTL mapping was carried out for protein-coding genes and for intergenic, multi-exonic lncRNA genes that were expressed

in at least one liver sample, namely: 20,559 (18,543 protein-coding and 2,016 lncRNA genes), 20,268 (18,275 protein-coding and 1,993 lncRNA genes), and 20,190 (18,207 protein-coding and 1,983 lncRNA genes) genes when analyzing all DO livers, male only DO livers, and female only DO livers, respectively. **Interpolating missing genotype information.** To preserve the extra information that the higher density SNP array provides, I performed QTL mapping on 64,713 SNPs markers (Table S3-2), which corresponds to the union of SNP markers from the two genotyping arrays used, after removing SNP probes with no differentiating information across the eight strains (Morgan et al. 2015). For SNP probes that were unique to one type of array, diplotype probabilities were assigned from the nearest measured SNP. All of the SNP probes that were analyzed are located in autosomes or on the X chromosome.

**Sex-biasedness of genetic association.** Sex-biased genetic associations for eQTLs that are significant in either male liver or female liver samples, are defined by LOD score calculated using male only DO liver samples - LOD score calculated using female only DO liver samples.

*Cis/trans* eQTLs. An eQTL was designated *cis* if the TAD where the gene is located overlaps at least one bp with the 95% Bayesian credible interval, which was defined based on the interval from the liver sample set (i.e. all livers, male only livers, or female only livers) that gives the most significant LOD. Each gene is assigned to the TAD where

its TSS occurs. TAD coordinates for mouse liver were based on (Matthews and Waxman 2017).

**eQTL overlap.** An eQTL overlaps a published eQTL if the gene name or RefSeq accession number matches, and if the associated variant/ associated region overlaps at least one bp the 95% Bayesian credible interval identified in this study for each eQTL. We used liftOver (Rosenbloom et al. 2015) to convert genomic regions into mm9 coordinates using default parameters. bioDBnet:db2db (Mudunuri et al. 2009) was used to convert Ensembl gene identifiers to RefSeq accession numbers, where applicable.

**Liver gene expression variation for sex-specific genes.** Liver expression variation for sex-specific genes were quantified across 112 male DO mice, and separately across 110 female DO mice fed on a standard chow diet using the var function in R.

**Co-regulated gene cluster.** Each eQTL region is centered at the SNP marker with the highest LOD score, i.e. the eQTL peak, boundaries are the  $2^{nd}$  SNP marker from the center. Overlapping eQTL regions by at least one bp with a common regulating strain are assigned as co-regulated clusters. Co-regulated gene clusters were identified based on eQTLs discovered in all DO liver samples, and separately male only DO liver samples, and female only DO liver samples. This analysis was based on eQTLs that were discovered for genes that are expressed (FPKM > 0) in at least 25% of samples in the

relevant liver samples sets, i.e., all DO livers, male only DO livers or female only DO livers.

**Gene expression correlation.** A batch effect that correlated with the GEO accession of the samples was removed from the gene expression level (in units of log2(FPKM+1)) using the ComBat function in the sva R package (Leek et al. 2012) with default parameters. The Pearson correlation was then used to calculate pairwise gene expression correlations.

**TF binding sites and DHS with strain-specific SNPs/indels at eQTL regions.** High quality SNPs/indels in the 8 DO mouse founder strains were downloaded from the Sanger Mouse Genome project (release 1211) (Keane et al. 2011). SNPs/indels that occur in only one founder strain were designated as strain-specific variants using a custom R-script provided in Supplemental file 1 (available upon request to <u>djw@bu.edu</u>). TF binding sites and DHS that are within each eQTL region (95% Bayesian Credible interval, defined based on the liver sample set that gives the highest LOD score), and that contain strain-specific SNPs/indels for the matching regulating strain, were selected for further analysis. For analysis depicted in Fig. 3-4, the following eQTL designations were used: eQTLs identified for genes with |male/female| > 4 were designated sex-specific (i.e., an eQTL for a sex-specific gene), and eQTLs identified for genes with |male/female| < 1.2 were designated sex-independent. Male/female expression ratios were determined by edgeR. eQTLs whose 95% Bayesian credible interval does not overlap by 1 bp with the TAD

that contains the gene it is associated with were designated *trans*. The number of eQTLs analyzed for the analysis of DHS and TF binding sites varied between 17 and 2,793, and is specified for each analysis in the x-axis labels of Fig. 3-4. For each eQTL region, we counted the fraction of DHS or TF binding sites that are male-specific, female-specific or sex-independent, as defined by their original studies, except for the set of ~72,000 liver DHS from (Sugathan and Waxman 2013) of which 4,644 DHS were designated male-specific and 2,814 were designated female-specific (Ling et al. 2010, Zhang et al. 2012, Conforto et al. 2015).

**Principal component analysis (PCA).** PCA was done using the prcomp function in R using log2, centered normalized microarray intensity values.

**DAVID enrichment.** DAVID functional annotation tool (<u>https://david.ncifcrf.gov/</u>) (Huang da et al. 2009, Huang da et al. 2009) was used to discover enriched KEGG pathways by providing the official gene symbol, i.e. gene name, of sex-specific proteincoding genes discovered in each strain. Pathways with adjusted P (Benjamini) < 0.05 were deemed significant.

# **3.D Results**

**Sex-specific gene expression patterns show high variability among DO mouse founder strains.** We identified 1,033 protein-coding genes that show sex-biased expression in mouse liver in one or more of the nine strains examined (418 male-biased genes and 615 female-biased genes) (male/female expression |fold-change| >1.5 based on microarray analysis; or male/female expression |fold-change| >2 based on RNA-seq, both at FDR < 0.05) (Table S3-1). These strains include CD-1 mice, used in our earlier studies of sex-biased liver gene expression (Lau-Corona et al. 2017), and the eight founder strains used to establish the DO mouse population (A/J, C57BL/6J, 129S1/SvlmJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ WSB/EiJ) (Churchill et al. 2012, Svenson et al. 2012). To further delineate differences in sex-biased gene expression across DO founder mice, I examined the subset of 471 sex-biased liver protein-coding genes that were identified using expression data obtained from the same microarray platform for each of the eight founder strains. Sex-biased gene expression was highly variable across DO founder strains, and was lost for 451 out of 471 (95%) of liver protein-coding genes in one or more of the eight founder strains. High variability of gene expression was also seen for the 74 most highly sex-biased protein-coding genes (male/female expression |fold-change| > 4 at FDR < 0.05 in at least one founder strain), where 58 of the genes (78%) did not show sex-specific expression in at least one founder strain (see Fig. S3-2 for examples of variable sex-specificity across DO founder strains). 220 of the 471 genes showed sex-biased expression in at least 2 of the 8 strains. Only 20 genes showed sex-biased expression in all 8 founder strains (Table S3-1). Heat maps confirmed the high variability of sex-specific liver gene expression across DO founder mice (Fig. 3-1A), and further, indicated that a subset of genes that showed sex-biased expression in only one strain were as highly sex-biased as genes that are sex-specific in

multiple strains. The heat map also revealed that the genes that are sex-specific in most, or in all, founder strains tend to be the most sex-biased (P < 2.7e-4, Student's t-test). Further, analysis of the 220 sex-specific genes with the highest ratio in male liver for [gene expression variability across DO founder strains] / [gene expression variability within strains], revealed prominent patterns of expression that are highly selective for each strain, further confirming the striking variability of sex-biased gene expression in DO founder strain livers (Fig. 3-1B).

The number of sex-specific genes identified in each founder strain ranged from 77 genes (strain A/J) to 183 genes (strain PWK/PhJ; Fig. 3-1C, Table S3-1), highlighting the possibility of strain-specific differences in the regulation of sex-specific genes. A similar pattern was seen for the number of highly sex-specific genes (male/female |fold-change| > 4; Fig. S3-3A) across strains, where A/J and CAST/EiJ mice have the lowest number of sex-specific genes. To identify similarities between sex-specific genes identified in each founder strain, I performed PCA based on gene expression levels (Fig. 3-1D). The largest variance (PC1 in Fig. 3-1D) corresponds to the sex of each liver, as expected, whereas the second principal component clustered male, and separately, female samples based on their evolutionary distance (PC2 in Fig. 3-1D; Fig. S3-3B) (Threadgill et al. 2011). The six *Mus musculus domesticus* subspecies strains are closer together, as compared to the other subspecies strains, i.e. *Mus musculus castaneus* (strain CAST/EiJ) and *Mus musculus musculus* (strain PWK/PhJ). Further, among the *Mus musculus domesticus* 

strains, the lone, wild-derived WSB/EiJ strain is separated from the other five strains, recapitulating the published phylogenetic tree (Fig. S3-3B) (Threadgill et al. 2011).

Next, I investigated if the sex-specific protein genes identified in each strain are enriched for different pathways. DAVID analysis (Huang da et al. 2009, Huang da et al. 2009) identified 16 KEGG pathways that are enriched (adjusted P < 0.05) in at least one of the eight sets of sex-specific genes (Fig. 3-1E). The enriched pathways are broadly divided into two categories (see dendrogram along y-axis, Fig. 3-1E): the first 10 pathways listed are enriched in select strains, while the last 6 pathways listed are significantly enriched in nearly all strains. Many of the pathways enriched in select strains have lower significance. Exceptions include pyruvate metabolism, which is uniquely linked to WSB/EiJ livers; and biosynthesis of unsaturated fatty acids and metabolism of xenobiotics by cytochrome P450, which are most strongly associated with PWK/PhJ livers.

I further explored if the variability of sex-specific gene expression patterns across DO founder strains affects the major pathway of GH-dependent gene regulation, as indicated by the response to hypophysectomy (Hypox). Hypox involves surgical removal of the pituitary gland, which ablates secretion of GH, as well as secretion of all other pituitary-dependent hormones. Two distinct classes of sex-specific liver-expressed genes have been identified based on their responses to Hypox. Class 1 sex-specific genes are activated in the liver by the pituitary GH secretion profile in livers of the sex where they

are more highly expressed, whereas class 2 sex-specific genes are repressed in the liver by the GH secretion profile of the sex where they are less highly expressed. Consequently, following Hypox, class 1 genes are down regulated in the sex where they show higher expression, whereas class 2 genes are up regulated (de-repressed) in the sex where they show lower expression. Using published lists of the four Hypox response classes of genes (male-specific class 1 and 2 genes, and female-specific class 1 and 2 genes (Table S3 of (Connerney et al. 2017)), we examined the distribution of each sexbiased gene class in each DO founder mouse strain. No noticeable difference was seen in the proportion of sex-biased genes in each of the four GH-responsive gene sets, as compared to the distribution reported for CD-1 mice (Fig. 3-1F; first bar on the left). Thus, the major GH-dependent regulatory mechanisms for liver sex-biased gene expression captured by the class 1/class 2 classification is largely preserved across DO founder strains, despite variability in the expression of many individual sex-specific genes.

**Genetic regulation of liver-expressed genes.** I analyzed 438 DO mouse liver samples (Munger et al. 2014, Chick et al. 2016, Gu et al. 2016) to identify SNPs that are significantly associated with the gene expression patterns of all liver-expressed genes (eQTLs). Genome-wide eQTL analysis was performed across all liver samples, and separately, across all male livers (n=219) and across all female livers (n=219). I thus identified 10,325 significant autosomal eQTLs associated with protein-coding and lncRNA genes when considering all livers, 7,414 autosomal eQTLs when considering

male livers only, and 8,233 autosomal eQTLs when considering female livers only (Table 3-1). These eQTLs can be merged into a set of 12,886 autosomal eQTLs (Fig. 3-2A). An additional 4,392 eQTLs were found on the X chromosome (Table S3-3B), resulting in 17,278 eQTLs identified genome-wide (Table S3-3A). 82% of the all liver, 65% of the male liver only, and 95% of the female liver only eQTLs are associated with autosomes (Table S3-4A). 51% of the autosomal eQTLs identified here confirm previously published eQTLs in mouse liver (Gatti et al. 2010, Munger et al. 2014, Chick et al. 2016, Gatti et al. 2017, Tyler et al. 2017); the other 49% of eQTLs are novel (6,344 out of 12,886) and includemany novel autosomal eQTLs for liver-expressed lncRNAs (Melia et al. 2016). More eQTLs were identified in female liver than in male liver for both proteincoding genes and lncRNA genes (Table 3-1), suggesting genetic regulation is more prevalent in female liver. A similar observation was reported in BXD mouse liver (Gatti et al. 2010). 66% of autosomal eQTLs that are within the same TAD (Matthews and Waxman 2017) as the genes they are associated with, indicating proximal genetic regulations dominates the eQTLs (*cis* regulation; Table 3-1).

For each eQTL, the strain whose SNP is most likely responsible for the genetic association captured by the eQTL ('the regulating strain'), was determined by identifying the founder strain with the largest regression coefficient, in absolute value, at the SNP marker where the strongest association occurs, i.e. the SNP marker with the highest LOD score. 60% (6,158 of 10,325) of autosomal eQTLs discovered when considering all DO livers are associated with genetic variants in either CAST/EiJ or PWK/PhJ mice (Fig. 3-

2B). These strains have the largest number of SNPs and indels (Svenson et al. 2012) and are the most evolutionary diverged, as compared to the other founder strains (Fig. S3-3). The WSB/EiJ strain follows with 10% all of autosomal eQTLs, and the other five founder strains are each associated with ~6% of autosomal eQTLs. This observation indicates that the gene expression pattern is altered as genetic variants accumulate during evolution. A majority of the autosomal eQTLs (54-67%) are associated with up regulation of gene expression in the regulating strain (Fig. S3-4). Some eQTLs have contributions from multiple regulating strains, i.e., the genetic variant of interest is associated with changes in gene expression in multiple strains. 35.8% (3,706 of 10,325) of autosomal eQTLs discovered using all DO liver samples have at least one other founder strain whose regression coefficient is within 20% of the identified regulating strain, indicating at least two strains make substantial contributions to the change in expression pattern (Table S3-4A).

1,308 genes are associated with multiple eQTLs (2,779 eQTLs; Table S3-4), suggesting these genes are subject to multiple, distinct forms of regulation. One example is the female-specific *Sult3a1* gene, which is associated with an eQTL on chromosome 10 that is highly significant in male DO livers, but is not detected in female livers, and with a second, distinct eQTL, on chromosome 9, that is uniquely identified in female livers (Fig. 3-2C). CAST/EiJ is the regulating strain for the eQTL on chromosome 10, and is associated with elevated expression of *Sult3a1* in male liver (Fig. 3-2D, *left*; Fig. 3-2E, *left*). In contrast, NZO/HIltJ is the regulating strain for the chromosome 9 eQTL, and is

associated with a distinct regulatory mechanism, leading to decreased expression of *Sult3a1* in female liver (Fig. 3-2D, *right*; Fig. 3-2E, *left*). Both eQTLs reduce the sex difference in *Sult3a1* expression in their respective regulating founder strains (Fig. 3-2E, *arrows*), but by different mechanisms.

Genetic regulation of sex-specific genes. 987 of the liver autosomal eQTLs are associated with one of the 1,033 sex-specific protein-coding genes or the 168 sex-specific lncRNA genes (Fig. 3-3A, Table 3-1). Most of the eQTLs associated with sex-specific genes were discovered when all DO liver samples were included in the analysis, but 66 and 71 eQTLs, respectively, were discovered only when male livers or only when female livers were used. Further, many of the eQTLs discovered using both sets of DO livers showed a stronger genetic association in one sex, as compared to the other (Table S3-4B). This trend is evident when LOD scores determined using male DO livers are compared to female DO liver LOD scores for the set of strongly sex-specific genes (Male/Female |fold-change| > 4 (Fig. 3-3B). eQTLs that are stronger in male DO livers are biased to be associated with strongly male-specific genes, and eQTLs whose association is stronger in female liver tend to be associated with strongly female-specific genes (Fig. 3-3B). Further, eQTLs with the greatest LOD score differences in male vs female liver (top 10%; 1,060 of 10,597 eQTLs; 172 of 1,060 eQTLs are identified for sex-specific genes; Fig. 3-2) are enriched 5.1-fold for being associated with genes with matched-sexspecificity, as compared to genes with the opposite sex-specificity (P-value < 6.5e-07, Fisher's Exact test). 315 sex-specific genes, however, are not associated with any eQTL

(Table S3-4C), indicating that the expression of these genes are less variable, as compared to sex-specific genes with eQTLs, across male DO (P < 0.0016, Student's ttest) and female DO (P < 0.011, Student's t-test) mice. The average gene expression level for these two gene sets are similar (P > 0.4 in both sexes, Student's t-test). Pathway enrichment on this gene set showed an enrichment for the PPAR signaling.

Next, I examined whether eQTLs with a sex-bias in genetic association may explain the variability in sex-specific gene expression across DO founder strains. I considered four conditions where an eQTL may result in the loss of sex specific gene expression (Fig. 3-3C) and four conditions where an eQTL may enhance sex-specificity (Fig. 3-3D). These analyses are based on 360 eQTLs (Table S3-5) that are significant in either male or female liver samples, which were examined for correlations with any of the 471 sexspecific protein-coding genes for which we have complete gene expression data in male and female livers for the eight DO founder strains. Overall, these 360 eQTLs are associated with 68% (322 of 471) sex-specific protein-coding genes. 172 of these eQTLs have characteristics that fit the first four conditions, i.e., they are associated with the loss of sex-specificity in the regulating founder strain (categories 1-4) and 188 eQTLs fit the second four conditions, i.e., they are associated with a gain in sex-specificity in the regulating strain (categories 5-8) (Fig. 3-3C, Fig. 3-3D; examples in Fig. S3-6). Remarkably, for 134 of the 172 category 1-4 eQTLs (78%), sex-specificity of gene expression is lost in the respective regulating strain, but is retained in at least one nonregulating strain. The fraction of these eQTLs that show a loss in sex-specificity ranges

from 53.1% (category 2 eQTLs) to 91.7% (category 1 eQTLs). Presumably the impact of many category 2 eQTLs is not strong enough to cause a loss of sex-specificity. In the case of the 188 eQTLs in categories 5-8, where an eQTL may increase the sex-specificity of gene expression, only 47% of eQTLs (85 of 188, with up to 60.8% in category 5) are associated with genes that are sex-specific in the matching regulating strain, suggesting there may be another layer of gene regulation, or these analyses may be limited by the dynamic range of our microarrays and hence their ability to detect sex-specificity. Overall, 60.8% of the 360 eQTLs that we examined (219 eQTLs associated with 205 genes; Table S3-6) were associated either with the loss or gain of sex-specificity of sex-biased genes in their matching regulating strains, highlighting the extensive role of genetic regulation in increasing the variability of sex-specific gene expression across DO founder mouse strains.

Genomic regions identified in an eQTL region may regulate sex-specific gene expression in the contributing founder strain through strain-specific genetic variants at sex-specific regulatory (e.g., enhancer) regions. These regulatory regions include DNase I hypersensitive sites (DHS) that are sex-biased in mouse liver, i.e., are significantly more open in one sex, as compared to the other, and sex-biased binding sites for key GHregulated liver TFs (Ling et al. 2010, Zhang et al. 2012, Conforto et al. 2015). Strainspecific genetic variants at these sites may alter or abolish GH-regulated TF binding activity and, perhaps alter chromatin accessibility (DHS) thereby dysregulate the expression of sex-biased genes. To investigate this hypothesis, we analyzed sex-biased DHS and sex-biased TF binding sites within eQTL regions (i.e. the 95% Bayesian credible intervals) and that contain SNPs/indels specific to the identified regulating strain (Table S3-4). Only eQTL regions that are < 3MB in width (8,254 of all 12,886 autosomal eQTLs; 64%) were used in this analysis. Sex-biased DHS were enriched at eQTL regions associated with genes with matching sex-specificity, as compared to eQTLs associated for genes expressed with the opposite sex-bias or genes with no sex-specificity (Fig. 3-4; top left, first and second panel). The majority of DHS found in any eQTL region, both for sex-biased and sex-independent genes, are sex-independent DHS; these DHS showed a significant enrichment to be at eQTLs for sex-independent genes, as compared to sexspecific genes (Fig. 3-4; top left, third panel). A similar pattern of enrichment was found for sex-biased STAT5 binding sites (Fig. 3-4, top right). In contrast, male-biased binding sites for the transcriptional repressor BCL6, which preferentially represses femalespecific genes in male liver (Meyer et al. 2009, Zhang et al. 2012), showed significant enrichment at eQTL regions for female-specific genes, as compared to male-specific or sex-independent genes. Moreover, sex-independent BCL6 binding sites were significantly depleted at eQTL regions associated with female-specific genes (Fig. 3-4, bottom). Further, BCL6 binding sites showed an even greater enrichment for trans eQTL regions, as compared to *cis* eQTL regions (Fig. 3-4; *bottom right*, as compared to *bottom left*), validating the proposed regulatory role of these *trans* eQTL regions. Collectively, these patterns ofsex-biased DHS and TF binding site enrichments provide strong support for the proposed regulatory role of these eQTL regions.
Gene expression co-regulation inference based on eQTLs. We hypothesize that genes with overlapping eQTLs are likely to be co-regulated, presumably due to their shared regulatory regions. Any overlapping eQTLs with a common regulating strain is defined as a co-regulated gene cluster. 1,521 co-regulated gene clusters, comprised of 4,105 genes, were identified based on eQTLs discovered in all DO liver samples (Table S3-7A). Most of the co-regulated gene clusters are small in size, with 65.8% of all clusters contains two genes each (Table S3-7D). The lack of many large co-regulated gene clusters suggests that genetic variants in DO mice are generally not at master regulatory regions for liver-expressed genes, i.e., regions where a single SNP dysregulates many genes. Rather, genetic variation typically impacts a small number (<10) of liverexpressed genes, leading to the observed variability of gene expression across strains, but without gross alteration of liver gene expression patterns.. The identified gene clusters show evidence of co-regulation, as many of their members show high pairwise gene expression correlation (see examples in Fig. 3-5). 50% (762) of the co-regulated clusters contain sub-clusters, where the identified eQTL regions are associated with up regulation of some genes in the cluster and down regulation of other genes in the cluster. One example is the 21-gene cluster on chromosome 2, which is predicted to occur in the PWK/PhJ strain (Fig. 3-5A). The first eleven overlapping eQTLs in the top cluster in Figure 3-5A have both positive regression coefficients (i.e. higher expression of the eleven listed genes; Fig. S3-7) and are located on the same chromosome with the genes they are associated with, whereas the last ten overlapping eQTLs have both negative regression coefficients (i.e. lower expression of the ten genes; Fig. S3-7 and Fig. S3-8)

and are located on a different chromosome with the genes they are associated with. Pairwise gene expression correlation analysis separates these eQTLs into two welldefined sub-clusters, where each sub-group has either a positive or a negative regression coefficient (Fig. 3-5A). The pairwise gene correlations are less tight in the negatively correlated sub-cluster (e.g. *Zbtb7c*), which suggests the eQTL association may result from indirect regulation or partial regulation involving another factor. Notably, in the case of the eQTL involving *Zbtb7c*, another, stronger genetic association was identified in *cis* (chr18, Fig. S3-8). The co-regulated cluster in Fig. 3-5A shows that genes from different chromosomes may be grouped together (Fig. S3-7), highlighting the ability of this approach to identify complex *trans* regulation. Overall, 10.4% (159) of co-regulated clusters group genes from multiple chromosomes. Further, 22.6% (344) of co-regulated clusters contain both protein-coding and lncRNA genes, generating testable hypothesis for possible functions of the clustered lncRNAs.

Co-regulated gene cluster based on overlapping eQTLs were also discovered in male only livers, and separately, female only livers. Differences in co-regulated gene clusters identified in male liver, and separately, in female liver were assessed for eQTLs with sexbiased genetic association. A limited number of co-regulated gene clusters show a switch in gene members in clusters identified in male liver samples, as compared to clusters identified in female liver samples. This is exemplified by three female-specific genes from the *Sulfotransferase* gene family: *Sult2a1*, *Sult2a3* and *Sult2a5* (Fig. 3-6A). In PWK/PhJ male mouse liver, but not in PWK/PhJ female mouse liver, *Sult2a1* and

Sult2a5 are predicted to be co-regulated, as their eQTLs are overlapping and are associated with the high expression of both genes (Fig. 3-6B; top left and bottom left). In contrast, in CAST/EiJ female liver only, Sult2a1 shows co-regulation with Sult2a3, and the high expression of both genes is associated with overlapping eQTL regions (Fig. 3-6B; top right and bottom right). Even though all four eQTL regions overlap, their eQTL peaks do not overlap, indicating a need for higher resolution genetic mapping to pinpoint the precise regulatory regions and tease apart these two distinct regulations. Sult2a5 and *Sult2a3* show no significant eQTLs in female or male liver samples, respectively (Fig. S3-9). The pairwise gene correlations between the three genes are consistent with these patterns, as Sult2a1 and Sult2a3 show higher correlation in female liver samples, as compared to male liver samples (Pearson correlation 0.7 vs. 0.3; Fig. 3-6A), and Sult2a1 and Sult2a5 show higher correlation in the male samples, as compared to female samples (Pearson correlation 0.8 vs 0.6; Fig. 3-6A). The relatively strong gene expression correlation in female livers between Sult2a1 and Sult2a5, which are not in the same CAST/EiJ co-regulated gene cluster, indicates that there is, nevertheless, an overall correlation in their expression across female DO mouse livers, once the complexity of other DO founder strains is factored in

#### **3.E Discussion**

Sex-specific gene expression characterizes many mammalian tissues (Rinn and Snyder 2005, Lowe et al. 2015), including liver (Zhang et al. 2011, Conforto and Waxman 2012, Krebs et al. 2012, Melia et al. 2016), where sex-biased genes are implicated in male-

female differences related to metabolism of drugs (Thurmann and Hompesch 1998, Waxman and Holloway 2009, Zanger and Schwab 2013) and xenobiotics (Mugford and Kedderis 1998), as well as disease susceptibility (Shimizu et al. 2007, Tessari et al. 2009, Baik et al. 2011). Pituitary GH has long been known to be a major regulator of hepatic sex-specific genes (Waxman and O'Connor 2006, Lichanska and Waters 2008) via its sex-dependent secretion pattern: pulsatile in male and near continuous in female (Jansson et al. 1985, Veldhuis et al. 2001), which in turn activate a cascade of TFs, including STAT5, HNF6 and BCL6 (Conforto et al. 2012, Zhang et al. 2012, Conforto et al. 2015). Here, we describe the role of genetic variants as another layer of regulation that works in tandem with the hormonal environment to control transcription of sex-biased genes. We used the DO mouse model, a genetically diverse population derived from eight inbred mouse strains (Churchill et al. 2012), to dissect the genetic component of sex-biased transcriptional regulation. We found that DO mice show high variability in their sexspecific gene expression patterns across DO founder strains, where the majority of sexspecific genes either loose or gain sex-biased expression in one or more founder strains. ~13,000 significant associations between autosomal genetic variants and gene expression (eQTLs) were identified in DO mouse liver, of which 49% are novel, and many are associated with lncRNAs. 987 of these eQTLs are associated with sex-specific genes, and a subset show stronger genetic association in one sex, as compared to the other. eQTLs with a large difference in genetic association between males and females (top 10%) are significantly enriched for an association with strongly sex-specific genes with matching sex-bias, as compared to genes of the opposite sex-specificity. 60.8% of eQTLs that are

associated with sex-specific genes and show a sex-biased genetic association are intimately linked with the loss or gain of sex-biased gene expression in the predicted regulating strain, highlighting the surprisingly extensive role of genetic variants in sexbiased gene expression. Analysis of the subset of TF binding sites and open chromatin regions (DHS) that contain genetic variants specific to the strain predicted to be regulated by eQTLs, revealed a significant enrichment of sex-specific STAT5 binding sites and DHS for eQTLs that are associated with genes with matching sex-specificity. Correspondingly, the male-biased transcriptional repressor BCL6 was enriched at eQTL regions mapping to female-biased genes, a pattern that was amplified in *trans* eQTL regions. Collectively, these patterns of enrichment highlight the regulatory nature of these eQTL regions. Finally, overlapping eQTLs were used to discover co-regulated gene clusters, some of which grouped genes from different chromosomes or grouped proteincoding genes with non-coding genes. These gene clusters provided evidence of coregulation based on high correlation of gene expression amongst cluster members.

Genetic factors promote variability of sex-biased gene expression across DO founder strains. Sex-specific gene expression shows exceptional high variability across DO founder mouse strains, where up to 95% of sex-specific genes either gain or loose sexbiased expression in at least one founder strain. Their high variability across strains indicates that they are not housekeeping or essential genes, however, they may contribute to clinically important variables, including liver metabolism and liver disease susceptibility. Further, the overall number of sex-specific genes varies more than 2-fold between strains. Genetic variability can increase or decrease the number of sex-specific genes, as the two strains that are the most evolutionarily divergent from the reference strain (C57BL/6J) have either the most sex-specific genes (PWK/PhJ) or the second fewest (CAST/EiJ). Remarkably, we demonstrated a tight connection between genetic variants and the variability of sex-specific gene expression across DO founder strains insofar as sex-specific gene expression was lost for 78% (134 of 172) of eQTLs which were associated with a decrease in sex-specificity in the regulating founder strains. Similarly, 47% (85 of 188) of eQTLs that were expected to increase the sex-specificity of gene expression were corroborated by the gain of the sex-biased gene expression in the respective regulating founder strains. Thus, genetic variants have an extensive impact in increasing sex-biased expression variability. The lower accuracy of our prediction for eOTLs that are predicted to increase sex-biased gene expression, as compared to those that ablate sex-specific gene expression (Fig. 3-3C, 3-3D) suggests another layer of gene regulation is involved; alternatively, the limits of the microarray data may decrease our ability to detect weaker sex-specific gene expression. Overall, 219 eQTLs were linked to the loss or gain of sex-specific expression of 205 genes (46% of the 471 sex-specific genes, for which we have complete expression in male and female liver of the 8 DO founder strains). The pervasive impact of genetic regulation on sex-biased gene expression is potentially greater, as up to 68% (322 genes; 360 eQTLs) of sex-specific genes are associated with eQTLs that were predicted to modulate the sex-specific gene expression across DO founder strains. Differences of the identified sex-specific genes across founder strains manifest in the significant enrichment of different KEGG pathways for select founder strains, most notably pyruvate metabolism for WSB/EiJ mice, and biosynthesis of unsaturated fatty acids and metabolism of xenobiotics by cytochrome P450 for PWK/PhJ mice. Further, DO founder mice exhibit diverse liver phenotypes (Svenson et al. 2007, Churchill et al. 2012), including differences in disease susceptibility, such as fatty liver disease (Fengler et al. 2016), liver fibrosis (Hillebrandt et al. 2005) and hepatoxicity (Tryndyak et al. 2012, Church et al. 2015), as well as differences in liver inflammatory responses (Bavia et al. 2015), whose molecular mechanisms are thought to involve sex-biased genes (Gonzalez 2005, Shimizu et al. 2007, Tessari et al. 2009, Ayonrinde et al. 2011, Baik et al. 2011, Takahashi et al. 2014).

We were not able to identify eQTLs for 315 sex-specific genes, and found their gene expression across strains to be less variable, thereby preventing us to uncover their regulatory regions based on association between genotype and gene expression data alone. These genes may be comprised of a core gene set whose sex-biasedness are essential to maintain, such that their gene expression variability across strains would not be tolerated. This gene set, notably, include important GH-regulated TFs including *Bcl6* and *Cux2*. A possible hypothesis is that only a limited number of sex-biased genes needed to establish the sexual dimorphism characterizing liver transcriptome and liver epigenome, and other non-essential sex-biased genes are allowed to vary, leading to diverse liver phenotypes.

Relationship between genetic variants and GH in the regulation of sex-biased gene expression. We show that the impact of a subset of genetic variants is substantially stronger in one sex, as compared to the other, despite the identical autosomal sequences of males and females. Thus, a mechanism must exist to allow genetic variants to confer regulatory control in a sex-biased manner. GH and its sex-dependent pattern of pituitary secretion is a key regulator of liver sex-biased gene expression, and is a primary candidate for this role. Differences in plasma GH profiles activate STAT5 in a sexdifferential manner, and STAT5, in turn, cooperates with other GH-dependent TFs to regulate the transcription of sex-biased genes through their sex-differential binding to liver chromatin. The sex-dependent plasma GH profiles are also linked to the widespread differences in histone modifications and chromatin accessibility between male and female liver (Ling et al. 2010, Sugathan and Waxman 2013). An earlier study on human whole blood samples found that the genetic control on both sexes are largely similar (Kassam et al. 2016), in contrast with a subset of eQTLs in our study showing strong genetic association selectively in one sex. This discrepancy may be due to the pervasive action of GH in imparting sex differences in regulatory regions in mouse liver (Ling et al. 2010, Sugathan and Waxman 2013, Connerney et al. 2017), which may be missing in human whole blood cells.

GH signaling pathways and GH-dependent sex-specific gene transcription seem to be largely unperturbed across DO founder strains, as indicated by the similar fraction of genes in each strain that belong to each of the four gene categories that are responsive to the loss of GH signaling in hypox mice, defined previously (Connerney et al. 2017). Further, eQTL regions, including a subset of trans eQTL regions, were enriched for sexbiased and GH-regulated sites of chromatin accessibility (DHS) and TF binding that match the sex-specificity of the gene that each eQTL is associated with, consistent with these eQTL regions playing a regulatory role. This analysis was based on the subset of sex-biased DHS and TF binding sites that encompass genetic variants specific to the founder strain predicted to be regulated by the respective eQTLs. Thus, certain genetic variants are proposed to be within GH-regulated accessible chromatins (DHS) and/or functional regulatory regions (TF binding sites) that are active in one sex, but not the other, which effectively creates a way for genetic variants to alter GH-regulated gene expression in a sex-biased manner. One example is the female-specific gene Cyp2b9, which loses its sex-biased expression in 129S1/SvImJ mouse liver. An eQTL with a strong repression effect (Fig. S3-6, *category 4*) uniquely linked to 129S1/SvImJ female mouse liver is characterized by 129S1/SvImJ-specific SNPs/indels at one female-biased DHS and 3 female-biased STAT5 binding sites. The genetic variants at one or more of these regulatory sites are proposed to interfere with the transcription of Cyp2b9 (Fig. S3-10A), leading to the dramatic decrease of Cyp2b9 expression seen in 129S1/SvImJ female liver (Fig. S3-10B and S3-10C).

**Properties of the genetic regulation of sex-biased genes**. The present study reveals several features that elucidate how genetic modifiers impact the transcription of sex-specific genes, including their tendency to regulate expression a small set of genes, rather

than via a master regulator susceptible to the effects of genetic variants. This conclusion is supported by our finding that co-regulated gene clusters, identified through overlapping eQTLs in the same regulating strain, are mostly small in size, i.e., clusters of <10 genes. However, this does not exclude the possibility that some of the genes regulated by genetic variants are themselves regulators, which perhaps regulate many genes, but which would not necessarily be sufficiently strong to be discovered as eQTLs in the present analyses. One such example is the *Rsl* gene, which codes for a Krüppel-associated box zinc finger (KRAB-ZPF) protein, whose expression varies across mouse strains (Krebs et al. 2003, Krebs et al. 2005). Rsl directly represses several sex-specific genes, including *Slp* and others, and it appears to regulate other sex-specific genes by indirect activation or repression (Krebs et al. 2009, Krebs et al. 2012). Thus, *Rsl* exemplifies how genetic variation can alter the expression of a sex-biased gene, which then causes a transcriptional cascade that has a widespread effect on sex-specific gene expression patterns. Another feature of genetic regulation in liver is a slight bias towards up regulation of gene expression, a property that is retained for strongly sex-specific genes. Finally, we found that eQTLs associated with highly male-specific genes tend to have stronger genetic regulation in male liver, and correspondingly, highly female-specific genes tend to have stronger genetic association in female liver.

In conclusion, we identified significant genome-wide associations between genetic variants and gene expression (eQTLs) in DO mouse liver, a subset of which showed stronger genetic association in one sex, as compared to the other. We further

characterized the high variability of sex-biased gene expression across DO founder strains, and showed that eQTLs with a sex-bias in association can explain the loss or gain of sex-specific gene expression, in DO founder strains, for 46% of sex-biased genes. eQTL regions associated with sex-specific genes, including a subset of *trans* eQTL regions, were shown to be enriched for sex-biased open chromatin regions and sex-biased binding sites for GH-regulated TFs. Finally, co-regulated gene clusters were identified based on overlapping regulatory regions (eQTLs), where a subset of clusters grouped highly correlated genes from different chromosomes. These findings reveal the extensive regulatory role played by genetic variants, working in tandem with GH, in the transcriptional regulation of sex-biased genes, many of which are implicated in the sex differential outcome of liver metabolism and disease susceptibility.

## Table 3-1. Number of autosomal eQTLs found when using all, male only or female only samples.

We observed more significant eQTLs in female than male. A similar observation was made in BXD liver, where 26.7% more significant eQTLs was reported for female liver as compared to male liver (1,638 significant eQTLs in male liver vs. 2,076 significant eQTLs in female liver).

						Combined
						(%in same
Types of	Set of	All	Male	Female		TAD
eQTLs	genes	Livers	Livers	Livers	Combined	(cis))
All	Protein-	9,088	6,494	7,219	11,362	
	coding (n =	(8,753	(6,202	(6,904	(10,097	7,492
	18,543)	genes)	genes)	genes)	genes)	(66 %)
		1,237	920	1,014	1,524	
	LncRNA	(1,177	(885	(957	(1,318	1,014
	(n = 2,016)	genes)	genes)	genes)	genes)	(67%)
	Total	10,325	7,414	8,233	12,886	8,506
Associated with sex- specific genes	Protein-	708	555	602		
	coding (n =	(682	(523	(581	830	631
	1,033)	genes)	genes)	genes)	(743 genes)	(76%)
		131	99	105		
	LncRNA	(121	(95	(97	157	108
	(n = 168)	genes)	genes)	genes)	(130 genes)	(69%)
	Total	839	654	707	<b>98</b> 7	739

# Table S3-1 to Table S3-8.Supplementary tables for Chapter 3 are available upon request (djw@bu.edu)

#### Figure 3-1. Protein-coding gene sex-specificity across DO founder strains. (A)

Heatmap of the log2 male/female fold-change values across DO founder mouse livers for 471 protein-coding genes that are significantly sex-specific (male/female |fold-change| > 1.5 at FDR < 0.05) in at least one DO founder strain based on the microarray dataset. Black boxes highlight genes with stronger sex-specificity in select DO founder strains. (B) Gene expression level for 220 sex-specific genes showing the highest inter-strain variability, but the lowest intra-strain variability (see Methods), across a set of 64 male DO founder mouse livers. The map is clustered based on the summation of expression levels of each gene across the 8 indicated DO founder strain male livers, with the color intensity indicating the fraction of expression seen in each liver. The data shown are based on 8 livers for each DO founder strain. (C) Number of sex-specific protein-coding genes in each DO founder strain, based on microarray datasets. (D) PCA of the 96 microarray samples (6 samples/sex/strain) based on the expression level of 471 sexspecific genes in each DO founder strain. (E) KEGG pathways enriched (adjusted Benjamini P < 0.05) in the set of sex-specific protein-coding genes identified in each DO founder mouse strain. (F) Distribution of male-biased class 1 and class 2 genes, and of female-biased class 1 and class 2 genes across mouse strains. The first bar presents the distribution of sex-biased genes across the four Hypox response classes (Connerney et al. 2017) identified in CD-1 mice. The other 8 bars represent the corresponding distributions in each of the 8 DO founder strains



Figure 3-2. Autosomal eQTLs for liver-expressed genes. (A) Venn diagrams showing numbers of autosomal eQTLs discovered using all DO livers, or using male only or female only liver samples (Table 3-1) and that are associated with protein-coding genes (left) or lncRNA genes (right). Overall, 10,597 autosomal eQTLs are significant in either male liver or female liver. (B). Pie chart showing the realtive frequency with which each of the indicated DO founder strains has the largest regression coefficient (i.e., is the major regulating strain) for the set of 12,886 combined autosomal eQTLs. (C) Genomewide association of Sult3a1 in all (left), male only (middle) and female only (right) livers. The horizontal red line marks the P < 0.05 significance cutoff based on the permutation test. (D) Regression coefficients (top of each panel) and LOD scores (log10(p-value); bottom of each panel) across the chromosome that have a significant eQTL peak in male (left) or female (right) mouse liver, as marked at bottom. Shaded area in the LOD score plot indicates the 95% Bayesian credible interval for each eOTL. (E) Expression of Sult3a1 in male and female mouse livers for each DO founder strain (left). Also shown are calculated male/female expression ratios (right). The middle hinge of each boxplot indicates the median, while the lower and upper hinges correspond to the first and third quartiles. Whiskers mark 1.5 \* IQR value. Red arrows indicate a decrease in female expression in NZO/HiLtJ mice and an increase in male expression in CAST/EiJ mice, both of which abolish the sex-specific expression seen in the respective mice strain.



**Figure 3-3.** Autosomal eQTLs associated with sex-specific genes. (A) Venn diagrams showing overlaps for eQTLs discovered in all, male only or female only liver samples that are associated with sex-specific protein-coding or sex-specific, multi-exonic intergenic lncRNA genes. (B) Distribution of LOD score differences (log10 values) in male vs. female liver for autosomal eQTLs that are associated with sex-independent (*left*), strongly male-specific (male/female |fold-change| > 4; *middle*) or strongly female-specific genes (male/female |fold-change| > 4; *left*). (C) (*Left*) Percentage of eQTLs in categories #1-4 (described at the right) that are associated with a loss of sex-specific gene expression in its regulating strain. (*Right*) Four possible scenarios, whereby an eQTL may result in the loss of sex-specific liver gene expression. (D) (*Left*) Percentage of eQTLs in categories #5-8 (described at the right) that are associated with a gain of sex-specific gene expression in its regulating strain. (*Right*) Four possible scenarios, whereby an eQTL may result in an increase in the sex-specificity of liver gene expression.



Fig. 3-3

Figure 3-4. Enrichment of DHS and TF binding sites that contain regulating strainspecific SNPs/indels at different types of eQTL regions. Fraction of male-specific, female-specific and sex-independent DHS (top left), STAT5 (top right), and BCL6 (bottom) binding sites with regulating strain-specific SNPs/indels at different sets of eQTLs. The different sets of eQTL are as follows: eQTLs identified for genes with |male/female| > 4 fold-change were designated sex-specific, eQTLs identified for genes with with |male/female| < 1.2 fold-change were designated sex-independent, and eQTLs identified in a different TAD from the gene it is associated with were designated *trans*. The number of eQTLs analyzed in each boxplot varied between DHS and TFs, and are specified in the x-axis labels. The middle hinge of each boxplot indicates the median value, while the lower and upper hinges correspond to the first and third quartiles. Whiskers marks 1.5 \* IQR value. P-values were determined by the Wilcoxon test, where \*\*\*=0.001, \*\*=0.01 and \*=0.05. The number of binding sites and DHS for each eQTL is shown in Table S3-4.





#### Figure 3-5. Co-regulated gene clusters based on overlapping eQTL regions

**discovered in all liver samples.** (A) A co-regulated cluster consisting of twenty-one overlapping eQTLs with the PWK/PhJ as the regulating strain. (B) Four co-regulated clusters containing sex-specific genes. The heatmaps show correlations of gene expression (Pearson)..







Α.





# **Figure 3-6. Co-regulated gene clusters that show different gene members in male vs. female liver.** (<u>A</u>) Pearson gene expression correlation of *Sult2a1*, *Sult2a3* and *Sult2a5* in male liver (*left*) and female liver (*right*). (<u>B</u>) Regression coefficients (top figure in each panel) and LOD scores (bottom figure in each panel) across chromosome 7 for *Sult2a1* in male liver (*top left*) and female liver (*top right*), for *Sult2a3* in male liver (*middle left*) and female liver (*top right*), for *Sult2a3* in male liver (*middle right*) and for *Sult2a5* in male liver (*bottom left*) and female liver (*bottom right*). Shaded area in the LOD score figures indicate 95% Bayesian credible interval for each eQTL.



Figure S3-1. Schematic overview of how the SNP array and RNA-Seq data were analyzed for each DO mouse liver samples.



**Figure S3-2. High variability of the sex-specificity of gene expression in individual livers of CD-1 mice and DO mouse founder strains for select male-specific and female-specific genes.** Gene expression in individual mouse livers across strains: male and female CD-1 mice (first row; left), male and female C57BL/6J mice (first row, right), male and female DO mice fed a standard chow diet (second row), or fed a high fat diet (third row). The fourth row shows box plots of gene expression level (FPKM) based on 128 individual male livers for DO founder mice fed a standard chow diet (left), or fed a high fat diet (right), and for male and female DO founder strain mice (fifth row; left). The first four rows are gene expression measured by RNA-Seq (FPKM values), while the gene expression on the fifth row was measured by microarrays. Male/Female sex ratios across the DO founder strains based on the microarray dataset was also calculated (fifth row; right). Examples shown are for a male-specific gene (*Cyp4a12b*), where male-biased expression is reduced or lost in PWK/PhJ mice, and a female-specific gene *Cyp2c39*, where female-biased expression is lost in C57BL/6J mice.

Fig. S3-2



119

Fig. S3-2 (contd)



#### Figure S3-3. Sex-specific protein-coding genes across DO founder strains. (A)

Number of highly sex-specific protein-coding genes (male/female |fold-change| > 4 at FDR < 0.05 in at least one DO mouse founder strain based on the microarray dataset) in each of the DO mouse founder strains. (B) Published phylogenetic tree (Threadgill et al. 2011) of the eight DO founder strains based on the SNP data for chromosome 11. The three DO founder strains whose names are in colored boxes are wild-derived inbred strains.



#### Figure S3-4. Properties of autosomal eQTLs in DO mouse liver.

Percentage of autosomal eQTLs that are associated with the up regulation or down regulation of gene expression in its respective regulating strain. There are 625, 672, 574, 593, 616, 3,085, 3,074 and 1,086 autosomal eQTLs whose regulating strain are either the A/J, C57BL/6J, 129S1/SvlmJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ, or WSB/EiJ strain respectively. This figure uses the 10,325 autosomal eQTLs found when all samples were used for eQTL discovery.

Fig. S3-4



### Figure S3-5. Examples of genes with stronger genetic association in male or female liver.

For each gene, the figure showsgene expression patterns across strains (see Fig. S3-2 for figure explanation), genome-wide eQTL scan (see Fig. 3-2C for figure explanation) and regression coefficients across the chromosome with a significant eQTL peak (see Fig. 3-2D for figure explanation). *Cav1* shows strong male-specific expression in the PWK/PhJ founder strain and in a subset of male DO mouse livers. Its strong eQTL is seen in male livers only and is associated with the PWK/PhJ strain. *Hsd3b5* shows male-biased expression in CD-1 mice and multiple DO founder strains; its expression in female liver is repressed in the C57BL/6J mice.

Fig. S3-5











Fig. S3-5 (contd)


# **Figure S3-6. Examples of sex-specific genes in categories 1-8, described in Fig. 3-3.** For each gene, the figure shows gene expression patterns across strains (see Fig. S3-2 for figure explanation), genome-wide eQTL scan (see Fig. 3-2C for figure explanation) and regression coefficients at the chromosome with a significant peak (see Fig. 3-2D for figure explanation).



Fig. S3-6 (contd)



Category 1 gene (regulated strain: CAST/EiJ; contd)



-2

NOD/ShiLtJ

NZO/HILtJ

CAST/EiJ PWK/PhJ WSB/EiJ

29S1/SvlmJ

C57BL/6J

ΓA

#### Category 2 gene (regulated strain: CAST/EiJ)

6

M L/A AU F 129S1/SvlmJ M 129S1/SvlmJ F NOD/ShiLtJ M NOD/ShiLtJ F

NZO/HILŁJ M NZO/HILtJ F CAST/EJ M

PWK/PhJ M PWK/PhJ F WSB/EiJ M

WSB/EJ F

CAST/EiJ F

C57BL/6J M C57BL/6J F Fig. S3-6 (contd)



## Category 2 gene (regulated strain: CAST/EiJ; contd)







Category 4 gene (regulated strain: 129S1/SvImJ)

6

(2951/SvimJ M (2951/SvimJ F NOD/ShiLtJ M NOD/ShiLtJ F

NZO/HILŁJ M NZO/HILŁJ F CAST/EJJ M CAST/EJ F PWK/PhJ M PWK/PhJ F WSB/EJ M WSB/EJ F 129S1/SvlmJ

NZO/HILtJ CAST/EjJ PWK/PhJ WSB/EjJ

C57BL/6J

٢N

C57BL/6J M C57BL/6J F

A/J M A/J F

Cyp2b9 expression across datasets CD1 mice (8 wks) C57BL6/J mice (15-20 wks) WY 200 100 NY 200 100 0 0 Female Individual mice samples Male Male Female Pooled samples DO mice (26 wks, standard chow) 800 600 WY400 200 200 0 Male Female Individual mice samples DO mice (26 wks, highfat chow) WY 400 200 0 Male Female Individual mice samples 8 DO founder strains (male, highfat chow) 8 DO founder strains (male, standard chow diet) 300 300 WY 200 100 WY42 100 0 0 NOD/ShiLtJ 29S1/SvlmJ NOD/ShiLtJ CAST/EiJ 129S1/Svlm. NZO/HILtJ C57BL/6J NZO/HILtJ PWK/PhJ C57BL/6J PWK/PhJ WSB/EiJ CAST/EiJ WSB/EiJ ٨J ٨J 8 DO founder strains (Microarray) Sex ratios (Microarray) 14 0 Log2(Intensity) 01 8 M/F Intensity Ratio -20 -40 -60

Fig. S3-6 (contd)

### Category 4 gene (regulated strain: 129S1/SvImJ; contd)





Fig. S3-6 (contd)

### Category 5 gene (regulated strain: PWK/PhJ ;contd)





140



## Category 6 gene (regulated strain: WSB/EiJ; contd)



## Category 7 gene (regulated strain: C57BL/6J; contd)





144



### Category 8 gene (regulated strain: PWK/PhJ; contd)

Figure S3-7. Regression coefficients (top figure in each panel) and LOD scores (bottom figure in each panel) for genes listed in Fig. 3-5A.



Fig. S3-7





Fig. S3-7(cont)



Fig. S3-7(cont)



Figure S3-8. Genome-wide LOD score plots for genes that are located in different chromosomes in the co-regulated cluster depicted in Fig. 3-5A.



Fig. S3-8





Figure S3-9. Genome-wide LOD score plots the Sult2a1, Sult2a3 and Sult2a5 genes.

## Figure S3-10. Strain-specific SNPs/indels are hypothesized to contribute to the loss of sex-specific gene expression for *Cyp2b9*.

(A) Hypothesis involving an eQTL with strong genetic association in female 129S1/SvlmJ mouse livers, discovered for female-specific gene *Cyp2b9*. In the identified eQTL region, we found one female-specific DHS and three female-specific STAT5 binding sites that contain 129S1/SvlmJ-specific SNPs/indels. These variants are hypothesized to interfere with the transcription of *Cyp2b9*, leading to a lost of sex-biased expression in 129S1/SvlmJ livers. (**B**) The distribution of gene expression level of *Cyp2b9* across individual DO mouse livers (Table S3-8) was used to discover eQTLs stratified by the genotype assigned at the the SNP marker with the highest LOD score. (**C**) A female-biased STAT5 binding site (fourth red arrow from the top) at a female-biased DHS (third red arrow from the top), containing two 129S1/SvlmJ-specific SNPs/indels (two top arrows), located within the eQTL region for *Cyp2b9* (circled).

## A. Cyp2b9 hypothesis

The 129S1/SvImJ-linked eQTL is significant in female liver, but not in male liver



Fig. S3-10



156

#### Chapter 4 Sex-biased gene co-expression networks in Diversity Outbred mouse liver

#### 4.A Abstract

Sex differences in liver gene expression and disease susceptibility are primarily regulated by pituitary growth hormone secretion patterns, which activate sex-dependent liver transcription factor, and establish sex-specific chromatin states. Ablation of pituitary hormone by hypophysectomy (hypox) identifies two major classes of liver sex-biased genes, defined by sex-dependent positive or negative responsiveness to hypox, respectively; however, the mechanism that leads to the different classes of hypox responsiveness is unknown. Here, we sought to identify candidate regulatory long noncoding RNAs (lncRNAs) that may control hypox responsiveness. First, we used mouse liver RNA-seq data for 30 different biological conditions to discover gene structures and expression patterns for  $\sim$ 15,500 mouse liver lncRNAs, including antisense and intragenic lncRNAs, as well as lncRNAs overlapping active enhancers, marked by enhancer RNAs. We identified 157 sex-specific lncRNAs that, were like many sexspecific protein-coding genes, regulated during postnatal liver development or were subject to circadian oscillations. Next, we utilized the high natural allelic variance of Diversity Outbred (DO) mice, a multi-parental outbred population, to discover tightly coexpressed clusters of sex-specific genes (gene modules) in male liver, and separately, in female liver. Sex differences in the gene modules identified were extensive. Remarkably, many gene modules were strongly enriched for male-specific or female-specific genes in one of the two hypox-response classes, indicating that the genetic heterogeneity of the

DO mouse captures responsiveness to hypox. Hypox-responsiveness was shown to be facilitated by multiple, distinct gene regulatory mechanisms, indicating its complex nature. Finally, we identified sex-specific lncRNAs whose expression across DO mouse livers showed an unexpected significant negative correlation with protein-coding gene modules enriched for genes of the opposite-sex bias and inverse hypox class, indicating strong negative regulatory potential for these lncRNAs. Thus, we used a genetically diverse mouse population to discover tightly co-expressed sex-specific gene modules that reveal broad characteristics of gene regulatory roles in controlling liver sex-bias gene expression.

#### **4.B Introduction**

Sex differences in hepatic gene expression are extensive (Waxman et al. 1985, Rinn and Snyder 2005, Clodfelter et al. 2006, Zhang et al. 2011), and are associated with sex differences in liver function and disease susceptibility (Shapiro et al. 1995, Tanaka 1999, Rogers et al. 2007, Baik et al. 2011, Li et al. 2012). These sex differences are primarily regulated by growth hormone (GH) (Cheung et al. 2006, Waxman and O'Connor 2006, Lichanska and Waters 2008), working through its sex-differential pituitary secretory patterns: pulsatile in male and near-continuous in female (Waxman et al. 1991, Veldhuis and Bowers 2003, Holloway et al. 2006, Conforto et al. 2012). The resultant sexdifferential plasma GH patterns regulate the activity of STAT5 and downstream transcription factors in a sex differential manner (Chia et al. 2006, Holloway et al. 2006, Conforto et al. 2012, Zhang et al. 2012, Conforto et al. 2015). These transcription factors are essential for the complex transcriptomic patterns specific to each sex (Clodfelter et al. 2006, Holloway et al. 2007) and help establish sex-differences in chromatin accessibility and chromatin states closely linked to sex-specific gene expression (Ling et al. 2010, Sugathan and Waxman 2013). The sex-biased transcriptome includes long noncoding RNAs (lncRNAs) (Melia et al. 2016), which have the potential for diverse regulatory roles, including epigenetic and transcriptional regulation of both in *cis* and in *trans* (Rinn et al. 2007, Huarte et al. 2010, Orom et al. 2010, Zhang et al. 2014, Amit-Avraham et al. 2015, Gayen et al. 2015).

Studies in rodent models have established that the prominent, sex-biased temporal patterns of GH secretion, are imprinted in males by neonatal exposure to androgens (Jansson et al. 1985). Sex-specific liver gene expression does not, however, become widespread until puberty (Conforto and Waxman 2012, Lowe et al. 2015). Surgical removal of the pituitary gland by hypophysectomy (hypox) identifies four major classes of sex-biased genes in mouse liver based on their responses to the loss of GH and other pituitary hormones (Wauthier et al. 2010, Connerney et al. 2017). Male class 1 and female class 1 liver sex-specific genes are activated by the GH secretion pattern of the sex where they are more highly expressed, whereas, male class 2 and female class 2 liver sex-specific genes are repressed by the GH secretion pattern of the sex where they are less highly expressed (Fig. S4-1). Consequently, following hypox, class 1 genes are down regulated and class 2 genes are de-repressed (up regulated) in mouse liver, which leads to

a near global loss of sex-specific gene expression (Wauthier et al. 2010, Connerney et al. 2017). The molecular bases of these four phenotypes in response to hypox is poorly understood.

Gene clustering has been used to identify groups of genes (gene modules/gene clusters) whose expression is strongly associated with different molecular phenotypes (gene signatures), including cancer subtypes (Alizadeh et al. 2000, Perou et al. 2000, Bhattacharjee et al. 2001, Sorlie et al. 2001) and various cellular processes (Ficklin et al. 2010, Eising et al. 2017, Tyler et al. 2017). An early study used gene modules discovered in mouse fat, brain, liver and muscle to study sex differences in the organization of gene expression networks. Gene modules discovered in liver and fat were found to be the least conserved between the sexes (van Nas et al. 2009).

Here, we use mouse liver RNA-Seq data for 30 distinct biological conditions to discover and characterize more than 15,000 liver-expressed lncRNAs, including many sexdependent lncRNAs subject to postnatal regulation during liver development (Peng et al. 2014, Lowe et al. 2015) or circadian oscillations (Menet et al. 2012, Fang et al. 2014). We exploited the high genetic variability of DO mice, a multi-parental outbred population derived from eight genetically divergent inbred mouse strains (Churchill et al. 2012, Schmidt 2015) to discover biologically meaningful clusters of sex-biased genes, and we investigated their enrichment for the above hypox-response classes. Our findings reveal a striking, and unexpected, significant negative correlation across a large panel of DO mouse livers between sex-specific lncRNAs and gene modules enriched for proteincoding genes of the opposite sex and inverse hypox-response class, generating testable hypotheses for lncRNA regulatory roles.

#### 4.C Materials and methods

LncRNA discovery. We collected 186 RNA-Seq samples of male and female mouse liver, representing 30 different biological conditions (Table S4-1A) (Menet et al. 2012, Renaud et al. 2014, Schmitt et al. 2014, Lowe et al. 2015, Connerney et al. 2017, Lau-Corona et al. 2017, Lodato et al. 2017). Tophat2 (Kim et al. 2013) was used to map a total of 10.7 billion reads to the C57Bl/6J mouse mm9 reference genome. LncRNA gene structures and isoforms were discovered using the method we recently published (Chapter 2) (Melia et al. 2016), with the following changes: (1) the expression filter removed lowly expressed genes at the level of 10% FPR, instead of 15%; (2) only gene structures overlapping one or more protein-coding gene exons on the same strand were removed. The latter modification allowed us to retain gene structures that correspond to antisense lncRNAs and intragenic lncRNAs. LncRNAs that do not overlap with any protein-coding gene were designated intergenic. LncRNAs that overlap protein-coding genes were designated antisense if they were transcribed on the opposite strand of the protein-coding gene; otherwise they were designated intragenic.

**Gene expression quantification for lncRNAs.** We devised four approaches for counting sequence reads to quantify lncRNA gene expression: (1) counting reads that overlap any

exon; (2) counting reads that overlap any region from the transcription start site (TSS) to the transcript end site (TES) (gene body); (3) counting reads that overlap genomic regions that were assigned as exons in all isoforms of a gene (exon-only counting); and (4) counting reads that overlap genomic regions that were designated introns in all isoforms of a gene (intron-only counting). These counting methods closely follow a recent publication for protein-coding genes (Connerney et al. 2017). Reads that mapped to regions that can be attributed to more than one gene were not counted, consequently, intragenic lncRNAs that are wholly within another gene will have no read counts when using the second or fourth counting method. All of the analyses presented in this study are based on the first counting method.

**Sex-specific and GH-regulated IncRNAs.** I used 8 mouse liver RNA-seq datasets (Table S4-1B) to assess the sex-specificity of lncRNA gene expression, as follows. Datasets # 1-3: Three independent poly(A)-selected total liver RNA datasets, encompassing 7 pools from each sex from CD-1 mouse liver (Connerney et al. 2017, Lau-Corona et al. 2017). Datasets # 4-6: For each sex, three pools of poly(A)-selected nuclear RNAs, three pools of ribosomal RNA-depleted total RNAs and three pools of ribosomal RNA-depleted nuclear RNAs from CD-1 mouse livers (Lau-Corona et al. 2017). Dataset # 7: 12 male and 12 female livers from 15-20 week old C57Bl/6J mouse liver (Lowe et al. 2015). Dataset # 8: 12 male and 12 female C57Bl/6J mouse liver at embryonic day 17.5 (Lowe et al. 2015). Differential analysis using edgeR (Robinson et al. 2010) was performed to compare male *vs* female livers in each of the eight datasets.

Any lncRNA with male/female |fold-change| > 4 at FDR < 0.05 in any of the eight datasets was designated as a sex-specific lncRNA. A listing of all 684 sex-specific lncRNAs and their expression data is provided in Table S4-2A. Pituitary regulation. Pituitary-regulated lncRNAs are those that show significant response to hypox in either male or female liver at hypox/intact |fold-change| > 2 and FDR < 0.05, as determined by edgeR. Sex-specific liver lncRNAs were categorized into four classes based on their response to hypox: class 1 and class 2 male-specific lncRNAs, and class 1 and class 2 female-specific lncRNAs. RNA-Seq data from two independent hypox mouse liver datasets was used to establish each lncRNA's response to hypox: PolyA-selected liver RNA and ribosomal-depleted liver RNA (Connerney et al. 2017). Analysis of the ribosomal-depleted hypox liver RNA dataset was carried out by Christine Nykyforchyn of the Waxman lab. All but four lncRNAs were assigned to the same hypox-response class by both hypox datasets; those four inconsistencies were resolved by choosing the classification indicated by the dataset with the higher male/female |fold-change|. The hypox-response class of each sex-specific lncRNA is shown in Table S4-2A.

**Regulated IncRNAs during liver development.** edgeR was used in ten separate differential expression analyses of C57BL/6J male mouse liver RNA-Seq datasets comparing lncRNA expression profiles between day 0 (immediately after birth and before the start of suckling) with liver RNA-Seq datasets for each of the following postnatal ages: day 1 (exactly 24 h after birth) and days 3, 5, 10, 15, 20, 25, 30, 45 and 60 (Robinson et al. 2010). LncRNAs with gene expression |fold-change| > 2 at FDR < 0.05
at any of the above ages were designated developmentally regulated lncRNAs. See Table S4-2A for developmental data for lncRNAs.

**Regulated IncRNAs in a 24-hour period.** Five differential expression analyses were performed to compare liver lncRNA RNA-Seq expression profiles in C57BL/6J male liver at Zeitgeber times (ZT) ZT6, ZT10, ZT14, ZT18, ZT22 to gene expression at ZT2 using edgeR. LncRNAs with gene expression |fold-change| > 2 at FDR < 0.05 at any ZT time point were designated circadian regulated lncRNAs. Circadian expression data for lncRNAs is shown in Table S4-2B.

Gene expression quantification in Diversity Outbred (DO) mouse liver. RNA-Seq data obtained for comprised of 112 male and 121 female DO mouse livers samples, all from mice fed standard chow diet, were downloaded from GEO, accession numbers GSE45684 (Munger et al. 2014, Gu et al. 2016) and GSE72759 (Chick et al. 2016). Read mapping and gene expression quantification for each sample was as described in Chapter 3. Briefly, reads were mapped to a specific diploid genome constructed for each individual mouse (Chapter 3). Gene expression was then quantified by counting sequenced reads that overlap any exon by at least one bp using featureCounts (Liao et al. 2014), using the best mapped location for each read. The restriction of only using unique reads for diploid genomes would limit the read counts to only include those reads that mapped to genome locations where the paternal and maternal allele differ, and thus could substantially underestimate the level of expression of any given gene. The gene expression level obtained for diploid genomes is based on the total number of reads that overlap the counted regions in either the paternal or maternal allele. Reads counts were then transformed to fragments per kilobase of exon per million reads mapped (FPKM) for downstream analysis, where gene lengths are the average length of the gene in the two alleles. A batch effect that correlated with the GEO accession of the samples was removed from the gene expression level (transformed to log2(FPKM + 1)) using the ComBat function in the sva R package (Leek et al. 2012). A listing of expression level across male and female DO mouse livers are in Table S4-7.

Gene expression quantification for antisense lncRNAs based on unstranded RNA-Seq datasets. The FDR for antisense lncRNAs was set to 1 for all differential analyses performed using unstranded RNA-Seq datasets, as the unstranded nature of such data prevents us from reliably attributing sequence reads to antisense lncRNAs. Unstranded RNA-Seq datasets include: the first 2 datasets used to determine lncRNAs sex-specificity (Table S4-1B; *first two rows*), polyA-selected hypox dataset used to identify pituitary hormone-regulated lncRNAs (Table S4-1A; *rows: male hypox and female hypox*), liver development RNA-Seq datasets in male liver (Table S4-1A; *any row with samples from* (Renaud et al. 2014)), RNA-Seq datasets for DO mouse liver samples, and RNA-Seq datasets that assess liver gene expression changes due to circadian rhythm (Table S4-1A; *rows: Male circadian ZT2-ZT22*). **Sex-specific and GH-regulated protein-coding genes.** We used the set of 1,033 sexspecific genes protein coding described in Chapter 3 (Table S3-1), which consists of genes that showed significant sex-specific gene expression in either CD-1 mouse liver or in any of the eight DO mouse founder strains (A/J, C57BL/6J, 129S1/SvlmJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ and WSB/EiJ) (Churchill et al. 2012, Svenson et al. 2012). A list of 531 pituitary hormone-regulated protein-coding genes, defined as genes responsive to hypox, was downloaded from Table S3 in (Connerney et al. 2017).

**Clustering sex-specific protein-coding genes.** I clustered sex-specific protein-coding genes based on their gene-expression correlation across 112 male, and separately across 121 female DO mouse liver samples using the weighted correlation network analysis (WGCNA) (Langfelder and Horvath 2008). WGCNA identifies co-expressed groups of genes (i.e., gene clusters/gene modules/gene networks) based on pairwise gene expression correlations across all samples. The correlation matrix is transformed into an adjacency matrix, which reflects the distance between genes, based on weighted gene expression correlations. Weights are introduced by raising the actual gene expression correlation to a power of  $\beta$ , which leads to more robust gene similarity measures. WGCNA parameters were set as follows: correlation function = the bicor function (Langfelder and Horvath 2012) with the maxPOutliers set to 0.1, type of network = signed hybrid, minimum module size = 5 genes, soft thresholding power ( $\beta$ ) = 3, deepSplit = 3 and cutTreeHeight = 0.25 to merge similar modules. Other parameters were

used at their default settings. Genes that were not expressed (FPKM = 0) in any set of livers (male liver samples, or separately, female liver samples) as well as those genes whose variance is 0 (MAD = 0) were excluded from the clustering analysis, this resulted in 1,018 and 1,014 of the 1,033 sex-specific protein-coding genes (Table S3-1) to be clustered in male and female DO liver samples, respectively; the overlap is high (1,013 of 1,019 genes) between the two gene sets.

**Relationship between protein-coding gene modules.** Similarities between proteincoding gene modules in each sex were measured by pairwise Pearson correlation between their first principal components, as calculated using the moduleEigenGenes function in WGCNA. The correlation matrix was then clustered using average linkage hierarchical clustering with default parameters.

Enrichment of various gene sets for protein-coding gene clusters (gene modules). To determine if any of the identified protein-coding gene clusters is enriched for any gene set with biological meaningful properties, I used the Hypergeometric test to assess the significance of overlap between the gene sets being compared; an overlap at p < 0.05 indicates significant enrichment for the overlapping gene set. Enrichment for each of the following gene sets was tested for each gene cluster: male-specific genes, female-specific genes, strongly sex-specific genes (i.e., male/female |fold-change| > 4), male class 1 genes, male class 2 genes, female class 1 genes, and female class 2 genes (see Table S3-1 for the first three categories, and see Table S3 in (Connerney et al. 2017) for members of

the hypox-response gene classes). To calculate the enrichment of male-specific genes in cluster *i*, we let mc = number of male-specific genes in cluster *i*, gc = number of genes included in the cluster analysis (1,018 sex-specific genes in male liver samples and 1,014 sex-specific genes in female liver samples, as described above) and m = the number of male-specific genes in g. Next, the probability of observing at least x male-specific genes in cluster i (P(#male-specific genes  $\geq x$ )) was defined by the probability of drawing x or more male-specific genes, given that we are drawing gc number of genes from g genes, which has m male-specific genes. This was implemented using the following R command: 1 - phyper(x-1, m, g-m, gc, lower.tail = TRUE, log.p = FALSE). The same formula was used to calculate enrichment for other gene sets, except we replaced the male-specific gene set with the other gene sets of interest.

**Properties of protein-coding gene clusters.** Nine properties were assessed for each gene cluster (see lower panels in Fig. 4-3 and Fig. 4-4): (1) the number of genes in each cluster; (2) the number of distinct TADs where the genes in each cluster are found; (3) the percentage of genes in each cluster with strong sex-specificity (male/female |fold-change| > 4 at FDR < 0.05; Table S3-1); the percentage of male-specific (4) or female-specific (5) genes in each cluster (Table S3-1); and the number of male-specific class 1 (6), male-specific class 2 (7), female-specific class 1 (8) and female-specific class 2 (9) genes in each cluster, expressed as a percentage of all hypox-responsive genes in each cluster (Table S3 in (Connerney et al. 2017)). The percentages in rows 6-9 of Fig. 4-3 and

Fig. 4-4 add up to 100%. For properties # 3-9, percentage values are only shown for those gene clusters with significant overlap with the relevant gene set, i.e. p-value < 0.05, as determined by the Hypergeometric test, described above.

**Correlating IncRNA with protein-coding genes clusters.** To identify IncRNAs that show a significant positive correction or negative correlation with protein-coding gene clusters, we calculated the Pearson correlation between: 1) the expression pattern of each sex-specific IncRNA across the full set of 112 male DO mouse livers, and separately, across the full set of 121 female DO mouse livers, and 2) the first principal component of each protein-coding gene module across the same DO mouse liver sets. The first principal component of each protein-coding gene module was calculated using the moduleEigenGenes function in WGCNA with default parameters. We determined all lncRNA-gene module correlation pairs in male samples, and separately in female samples. P-values for the significance of the correlations were obtained using the cor.test function in R and adjusted by the FDR method. Due to the unstrandedness of the DO mouse liver RNA-Seq datasets, this analysis was limited to the set of 168,multi-exonic, intergenic sex-specific lncRNAs (Table S4-2A). Top correlations were those that met a pvalue of < 0.01 at FDR < 0.001.

**eRNA identification.** 27 GRO-Seq C57BL/6J male liver samples were downloaded using GEO accession # GSE36916 and GSE59486 (Menet et al. 2012, Fang et al. 2014). Poly-A tails were trimmed from sequence reads using HOMER (Heinz et al. 2010)) with the following parameters: motif: AAAAAA or TTTTTT, mismatch = 2, minMatchLength = 5 and -min = 20. Sequence reads were mapped to the C57BL/6J mm9 reference genome using Bowtie2 (Langmead and Salzberg 2012), and mapped reads were used to discover peaks using HOMER with the follower parameter settings: L = 3 and bodyFold = 3. Every peak in the positive strand was paired with the nearest peak in the reverse strand. Any peak pairs that were > 500 bases apart and converging into the same direction were removed. eRNA loci were defined as the 1 kb region centered at the overlapping region in each peak pair. The 15,853 eRNA loci that overlap any published H3K27ac or H3K4me1 peaks for young adult mouse liver (Sugathan and Waxman 2013) were retained (Table S4-6).

#### 4.D Results

**Discovery of 15,558 liver-expressed lncRNAs.** We recently described 4,961 liverexpressed intergenic lncRNAs (Chapter 2) (Melia et al. 2016), and characterized their gene expression, conservation level and transcriptional regulation. We, and others, have since produced many strand-specific RNA-Seq datasets, where the direction of transcription is preserved. These datasets enabled us to discover other types of lncRNA that were not included in our initial study, namely antisense and intragenic lncRNAs, many of which may contribute to the transcriptional regulation of nearby genes (Zhang et al. 2014, Amit-Avraham et al. 2015, Schein et al. 2016). We collected 186 RNA-Seq datasets encompassing mouse liver representing 30 different biological states, spanning from embryonic day 15.5 to 15-20 week old mice (Fig. 4-1A). Other biological conditions represented include livers from each of the following: mice treated with foreign chemical agonists of liver nuclear receptors, mice whose GH secretion patterns were altered, and mice euthanized every 4 h to identify gene expression changes due to circadian rhythm (Menet et al. 2012, Renaud et al. 2014, Schmitt et al. 2014, Lowe et al. 2015, Connerney et al. 2017, Lau-Corona et al. 2017, Lodato et al. 2017). A total of 10.7 billion reads were mapped to the reference genome. LncRNAs were discovered using the approach described in Chapter 2 (Melia et al. 2016), except that we only removed gene structures whose exons overlap protein-coding gene exons transcribed in the same direction (Fig. S4-2). 15,558 lncRNAs were discovered (Table S4-2A), of which 13,343 do not overlap any protein-coding gene, and were therefore designated intergenic (Fig. 4-1B). 1,966 lncRNAs were designated antisense lncRNAs based on their overlap with the antisense strand of protein-coding genes (Fig. 4-1C). Finally, 249 lncRNAs were intragenic, i.e., transcribed in the same direction as an overlapping protein-coding gene, but without any exonic region overlap (Fig. 4-1D). 16% (2,098) of the intergenic lncRNAs have multiple exons (multi-exonic lncRNAs), while 47% (916) of the antisense and 55% (138) of the intragenic lncRNAs are multi-exonic. A subset of lncRNAs (18%; 2,847) may be involved in the activation of other genes due to their overlap with enhancer RNAs (eRNAs) (Table S4-2A), which were previously shown to correlate positively with enhancer activity (Kim et al. 2010). Overall, 44% of the 15,558 liverexpressed lncRNAs are not found in any prior annotation or database for noncoding RNAs (Melia et al. 2016, O'Leary et al. 2016, Zhao et al. 2016).

Regulation of sex-specific lncRNAs in male liver. The liver transcriptome undergoes significant changes during postnatal liver development (Peng et al. 2014, Gordillo et al. 2015). Notably, sex-biased gene expression emerges at  $\sim 4$  weeks of age, with many more sex-specific genes showing developmentally regulated expression in male liver as compared to female liver (Conforto and Waxman 2012). We identified 684 sex-specific IncRNAs that showed significant sex-biased gene expression (male/female |fold-change| > 4 at FDR < 0.05; Table S4-2A) in any of 8 independent datasets comparing male liver to female liver samples (Fig. S4-3, *left panel*). 168 of these lncRNAs were multi-exonic and intergenic. We assessed their regulation during postnatal male liver development, by comparing their gene expression patterns on day 0 (immediately after birth and before the start of suckling) to that on day 1 (i.e., 24 h after birth), and to that on days 3, 5, 10, 15, 20, 25, 30, 45 and 60 of age (Fig. S4-3, right panel). 144 of the sex-specific lncRNAs were regulated at least 2-fold at FDR < 0.05 on at least one of the above developmental time points (Fig. 4-2A; Table S4-2A and S4-2C). The majority of male-biased lncRNAs were up regulated by 25 days of age in male liver, consistent with the patterns found for male-biased protein-coding genes (Fig. 4-2, gene set I) (Conforto and Waxman 2012). A subset of the male-specific genes in set I, however, showed a significant increase in expression as early as the first week of life. Correspondingly, a subset of female-specific genes was down regulated in male liver by 20 days of age (Fig. 4-2, set II). A smaller subset of male-specific genes was strongly repressed after birth, and showed either continued repression (Fig. 4-2, set III) or up regulation around 30 days of age (Fig. 4-2, set IV), while maintaining their male-biased expression in adulthood. Most femalespecific genes were up regulated earlier in male liver development and were either maintained or tapered off their up regulation at adulthood (60 days of age; Fig. 4-2, *set V*). The latter gene set still showed female-biased expression in adulthood, likely due to a stronger up regulation in female liver, as compared to male liver. This pattern was previously seen for a limited number of female-specific protein-coding genes (Conforto and Waxman 2012).

Many organisms exhibit oscillating gene expression patterns with a period of  $\sim$ 24-hour (circadian rhythms). In mammals, these circadian patterns are maintained by feedback loop mechanisms orchestrated by several transcription factors (Ko and Takahashi 2006, Dardente and Cermakian 2007, Menet et al. 2012). We quantified gene expression for lncRNAs expressed in male liver, every 4 h in a 24-h period (ZT2, ZT6, ZT10, ZT14, ZT18 and ZT22). Five separate differential expression analyses were performed to compare ZT6 - ZT22 to ZT2, where ZT2 corresponds to 2 h after animals were exposed to light during a 12-h light and 12-h dark cycle. 241 lncRNAs, including 52 sex-specific lncRNAs, showed regulation at one or more circadian time points (|fold-change| > 2 at FDR < 0.05; Fig. 4-2B; Table S-42B and Table S4-2D). Subsets of both male-specific and female-specific lncRNAs showed higher expression at one or more time points after ZT2 (Fig. 4-2B, set I and III; Fig. 4-2C, top), and another subset showed the opposite pattern (Fig. 4-2B, set II and IV; Fig. 4-2C, bottom). We identified fewer circadian regulated female-specific lncRNAs, especially those that were down regulated after ZT2, most likely due to the exclusive use of male livers in this analysis. Overall, 157 sexbiased lncRNAs were regulated during postnatal liver development or responded to the circadian rhythm.

### Protein-coding gene clusters are enriched for biologically relevant traits. Sexspecific genes display at least two distinct phenotypes following the ablation of pituitary GH secretion by hypox, indicating there are at least two distinct classes of male-biased and two classes of female-biased genes, one subject to positive regulation and the other subject to negative regulation by GH (Wauthier et al. 2010, Connerney et al. 2017). Little is known about the underlying mechanisms whereby these distinct regulatory responses to hypox are achieved. We hypothesized that sex-specific lncRNAs might play a role in this regulation. To examine this possibility, I first examined whether a strong genetic regulatory component is evident for the four distinct hypox-response classes. To accomplish this, I utilized a large panel of liver gene expression datasets available for DO mice, which have a natural high allelic variance that manifests as a variable gene expression pattern in each individual mouse. Specifically, I used WGCNA (Langfelder and Horvath 2008) to cluster sex-specific protein-coding genes based on their gene expression correlations across male DO livers, and separately, female DO livers. I identified 40 gene clusters in male livers, and 44 other gene clusters in female livers (Fig. 4-3 and Fig. 4-4; Table S4-3A and Table S4-3B). The average number of genes per cluster was comparable in the two sets of livers (25.5 genes/cluster for male livers, 23 genes/cluster for female livers). Unexpectedly, more than half of the sex-biased gene clusters were significantly enriched for either male-specific or female-specific protein

coding genes (Fig. 4-3 and Fig. 4-4; rows 4-5), even though the analyses were performed using expression data from a single sex. Even more striking, many of the clusters were significantly enriched for one of four biologically relevant traits, defined by the change in gene expression following hypox, i.e., class 1 or class 2 male-specific genes, and class 1 or class 2 female-specific genes (Fig. 4-3 and Fig. 4-4; rows 6-9; Table S4-4A-D). These enrichments are unexpected, as the liver samples used to discover gene clusters come from DO mice, which have an intact pituitary gland, and are thus not expected to manifest the four hypox-responsive phenotypes. Further, in separate analyses, we determined that each of the 8 DO mouse founder strains is characterized by a similar distribution of sex-biased genes in each of the four hypox response classes (Chapter 3), indicating that a subset GH regulatory pathways controlling sex-biased liver gene expression are still present in the DO founder strains. Thus, the significant enrichment of many DO mouse sex-biased liver gene clusters for a single hypox-response class indicates that the genetic diversity of DO mice captures a key mechanistic characteristic of sex-biased gene regulation. A subset of gene regulation mechanisms related to hypoxresponsiveness is presumably altered by genetic variants accumulated in DO mouse founder strains.

Overall, in the DO male liver samples, 13 gene clusters encompassing 240 male-specific protein-coding genes were significantly enriched (p < 0.05, Hypergeometric test) for male-specific genes, and an additional 10 clusters encompassing 379 female-specific protein-coding genes were significantly enriched for female-specific genes (Table 4-1).

Further, a subset was significantly enriched for strongly sex-specific genes (|fold-change| > 4; Fig. 4-3 and Fig. 4-4; *row 3*). Significant enrichments were also found in DO male liver for the four hypox-response gene classes (Table 4-1): 7 gene clusters (encompassing 85 genes) were enriched for male class 1 genes, 2 gene clusters (encompassing 12 genes) were enriched for male class 2 genes, 3 clusters (encompassing 21 genes) were enriched for female class 1 genes, and 6 clusters (encompassing 89 genes) were enriched for female class 2 genes. 14 of 16 clusters that are enriched for one of the four classes of hypox-responsive genes are enriched for sex-specific genes. Protein-coding gene clusters with similar patterns of enrichment for sex-biased genes, and for hypox-response gene classes, were discovered in DO female liver samples, as is summarized in Fig. 4-4 and Table 4-1. The identified clusters do not consist of genes from the same genomic regions, as a majority of genes in every cluster are from different topologically associated domains (TADs) (Matthews and Waxman 2017) (Fig. 4-3 and Fig. 4-4; *row 2*), indicating the potential for this approach to identify *trans* regulation.

Although overall cluster statistics and enrichments are similar between male and female liver (Table 4-1), ~82% (33 of 40) of gene clusters discovered in male liver do not have any counterpart in the female liver clusters; only 7 of the 40 DO male liver clusters overlap (defined as >50% genes in common) with any cluster discovered in female liver (Fig. S4-4). Five of the 7 overlapping clusters (circled in Fig. 4-5) are enriched for either male-specific or female-specific genes, but not for highly sex-specific genes. An earlier study, using another mouse strain, found a higher overlap (53%) between gene clusters in the two sexes, which may be due to those analyses clustering all varying genes, as opposed to only sex-specific genes. The degree of gene expression correlation within a cluster varied among cluster members (Fig. S4-5), with some but not other genes showing strong correlations with many genes in the cluster. Thus, there is an inherent structure of gene regulation for each cluster, where genes with many connections to other genes may play an important role in regulation.

Sex-biased lncRNAs negatively correlate with sex-opposite and inverse hypox classenriched protein-coding gene clusters. Next, we considered whether any of the 168 multi-exonic, intergenic sex-specific lncRNAs identified above might serve as a regulator of any of the sex-specific protein-coding gene clusters (gene modules), as indicated by gene expression correlations. Significant correlations (FDR < 0.001) were seen between sex-biased protein coding gene modules and 73 sex-specific lncRNAs, 40 showing correlations in DO male liver (Table S4-5A) and 52 in DO female liver (Table S4-5B). Most of these were positive correlations, which could indicate co-expression/coregulation between the lncRNA and the protein-coding gene module. I therefore focused on negative regulators, because of the difficulty of distinguishing a lncRNA that serves as a positive regulator of a gene module from a lncRNA that is simply a co-expressed gene. I identified male class 1 lncRNAs whose expression across DO mouse livers shows a significant negative correlation with protein-coding gene clusters enriched for female class 2 genes in male liver, as well as female class 1 lncRNAs that are negatively correlated with male class 2-enriched protein-coding gene modules in female liver, and

correspondingly for class 2 sex-specific lncRNAs (Fig. 4-6). A negative correlation of gene expression profiles between inverse hypox classes of the opposite sex is consistent with the models shown in Fig. 4-6, and would enable a lncRNA to negatively regulate protein-coding genes by one of several established mechanisms (Halley et al. 2014, Gayen et al. 2015, Ohhata et al. 2015). 16 sex-biased multi-exonic lncRNAs showed significant (FDR < 0.001) negative correlations with gene clusters enriched (P < 0.01) with sex-opposite protein coding genes of the inverse hypox class (Table 4-2 and Table S4-8), consistent with the proposed negative regulatory role for these lncRNAs (c.f. Fig. 4-6). In some cases, the correlations involve male-specific lncRNAs whose expression in female DO liver negatively correlated with clusters enriched for highly female-specific protein-coding genes (e.g., ncRNA inter chr10 9000) or female-specific lncRNAs whose expression in male DO liver negatively correlated with clusters enriched for highly male-specific protein-coding genes (e.g., ncRNA inter chr2 1430) (Table 4-2). In these cases, low levels of the male-specific lncRNA would need to be maintained in female liver in order for the inverse hypox class, female-specific gene module to maintain its high expression in female liver, and correspondingly for female-specific lncRNAs and inverse hypox class, male-specific gene modules in male liver. Overall the negatively correlated sex-biased gene clusters shown in Table 4-2 encompass a substantial fraction (327 of 1,019) of sex-biased protein coding genes included in our clusters (182 malebiased genes and 145 female-biased genes; 75 male class 1, 11 male class 2, 22 female class 1 and 35 female class 2 genes). Example of individual clusters are shown in Fig. S4-5.

Negative correlations between sex-opposite lncRNAs and inverse hypox class-enriched protein-coding gene clusters (Table 4-2) were also seen when we included more weakly correlated (FDR < 0.05) lncRNAs, particularly in male liver (Fig. 4-7, *right* panels). Fig. 4-7A plots the hypox class distributions for 94 sex-specific lncRNAs that show positive (*left*) or negative correlations (*right*) with 16 protein-coding gene modules that are enriched for one of the four hypox-response classes. As expected, more male class 1 lncRNAs are positively correlated with gene modules with strong enrichment for male class 1 protein-coding genes, as compared to those that are not (Fig. 4-7A left; first eight bars vs the other bars), or as compared to the overall distribution for the 142 sex-specific lncRNAs that are varying in male liver (MAD > 0; Fig. 4-7A All; Table S4-2E). A similar pattern is seen in female DO livers (Fig. 4-7B left; first six bars vs the others, and Fig. 4-7B *left* vs *All*; Table S4-2F). Negative correlations between inverse hypoxresponse classes were also seen, with more female class 2 lncRNAs being negatively correlated with protein coding gene modules enriched for male class 1 protein-coding genes, as compared to those that are not (Fig. 4-7A right; first seven bars vs the others) or as compared to the overall distribution for all sex-specific lncRNAs that are varying in male liver (Fig. 4-7A right vs All; Table S4-2E). A similar, albeit less prominent, negative correlation between female class 1 lncRNAs and protein-coding gene modules enriched for male class 2 genes was observed (Fig. 4-7B right; first two bars vs the others). Taken together, these findings support a model whereby class 1 sex-biased

lncRNAs negatively regulate protein-coding modules that are enriched for class 2 sexbiased genes of the opposite sex-specificity, and vice versa.

#### 4.E Discussion

Many genes in mammalian liver are differentially expressed between males and females (Rinn and Snyder 2005, Clodfelter et al. 2006, Zhang et al. 2011, Lowe et al. 2015), including lncRNA genes (Melia et al. 2016), which are thought to contribute to liver sex differences by establishing or maintaining the widespread sex-differences in chromatin accessibility and chromatin states that are closely linked to sex differences in liver transcription factor binding and chromatin states (Ling et al. 2010, Zhang et al. 2012, Sugathan and Waxman 2013, Conforto et al. 2015). Here, we expanded the repertoire of liver-expressed ncRNAs to include more than 15,500 lncRNAs, including many novel genes, based on liver gene expression data collected under 30 distinct biological conditions. A subset of these lncRNAs showed strong sex-biased expression, as well as regulation during liver development or in response to circadian rhythm. Further, we identified sex-specific lncRNAs whose expression across DO mouse livers shows a significant negative correlation with protein-coding gene modules enriched for the genes of the opposite sex-bias and inverse hypox-response class, consistent with the proposal that these lncRNAs play an important negative regulatory role in controlling sex-biased gene expression. The broad characteristics of sex-biased gene expression revealed by this study include an indication that multiple, distinct gene regulatory mechanisms likely contribute to the four classes of hypox-responsive genes. These findings illustrate the

usefulness of gene modules for generating hypotheses for lncRNA function, and reveal underlying characteristics governing sex-biased gene regulation.

**Regulation of lncRNA gene expression in male liver**. We examined the expression of sex-biased lncRNAs during postnatal male liver development, where a majority of changes relevant to the emergence of liver sex-biased gene expression have been found to take place (Conforto and Waxman 2012). We found that a majority of male-specific lncRNAs were up regulated postnatally, as early as the first week of life. Further, we identified two patterns of up regulation for female-specific lncRNAs: sustained or dynamic up regulation after birth, with the latter pattern showing either down regulation or reduced up regulation of gene expression around 4 wk of age. These sex-specific IncRNAs may contribute to the acquisition of sex-biased gene expression at puberty and in early adulthood. We also evaluated the regulation of sex-specific lncRNAs by circadian rhythms in male liver, which gives rise to a diurnal rhythm affecting the expression of a large number of liver-expressed protein coding genes (Menet et al. 2012, Fang et al. 2014). Sex-biased lncRNAs showing significant gene expression changes in response to the alternating light and dark cycle in male liver could contribute to the regulation sex-specific protein-coding genes with the matching circadian rhythm, which characterizes a subset of liver sex-biased genes (Menet et al. 2012).

**Complex regulation characterizes the four hypox-response gene classes**. We identified protein-coding gene modules in male liver, and separately, in female liver, of

which almost half of the identified clusters are enriched for either male-specific or female-specific genes. Hierarchical clustering of these gene modules showed that, both in male and female DO mouse liver, almost all of the modules that are enriched for malespecific genes segregated into their own branch, separated from the gene modules that are enriched for female-specific genes (Fig. 4-3, Fig. 4-4). This indicates that genetic variation in the population has a major impact on sex-biased gene expression, which in the case of DO mice is easy to discern given the genetic variability in this mouse population. One mechanism for genetic factors to exert their regulatory effect in a sexbiased manner is through genetic variants that are within GH-dependent open chromatin regions (DHS) or GH-regulated TF binding sites. GH has been shown to impart sex differences in liver through its regulation of several GH-dependent transcription factors (Clodfelter et al. 2006, Clodfelter et al. 2007, Holloway et al. 2007, Conforto et al. 2012, Zhang et al. 2012, Conforto et al. 2015). On the other hand, gene modules that were enriched for one of the four hypox-response gene classes were found in multiple, separate branches, and mixture of both class 1 and class 2 genes can be found within the same branch (Fig. 4-3, Fig. 4-4). Each co-regulated gene cluster enriched for one of the four hypox-response classes is thus likely to be distinctly regulated, which in turn shapes their gene expression pattern differently across DO mice, as compared to other gene clusters enriched for the same hypox-response gene class. This suggests that multiple, distinct regulatory mechanisms are at play in shaping each of the four major classes of hypoxresponsive genes, highlighting the complexity of their regulation. Further characterization of the identified gene clusters may include discovery, for each gene cluster, of

enrichments for TF binding sites, developmental time points, or circadian groups, all of which may help delineate molecular mechanisms characterizing the four hypox-response classes. Protein-coding gene modules discovered in male and female liver showed substantial overlap for only 7 clusters, encompassing 223 (of 1,033) sex-specific genes based on clusters discovered in male liver, indicating that a large fraction of gene regulation governing sex-biased expression involves sex-differential mechanisms.

Sex-biased lncRNAs are proposed to negatively regulate select protein-coding gene modules. 73 multi-exonic, intergenic sex-specific lncRNAs showed a highly significant correlation (FDR < 0.001) with at least one sex-specific protein-coding gene module. Further, 16 of these sex-specific lncRNAs showed significant negative correlation with a protein-coding gene module(s) enriched for genes with the opposite sex-bias and inverse hypox-response class, allowing us to hypothesize that these lncRNAs negatively regulate their inversely correlated protein-coding gene modules. The eight scenarios, where negative gene expression correlations between inverse hypox classes may indicate negative regulatory role (Fig. 4-6), can be broadly divided into two groups: in one group, high expression of sex-specific lncRNAs is needed to repress protein-coding genes of the opposite sex-bias, e.g., male class 1 lncRNAs and female class 2 gene modules in male liver; and in the second group, lower expression of lncRNAs is needed to prevent repression of protein-coding genes of the opposite sex bias, e.g., male class 1 lncRNAs and female class 2 gene modules in female liver. These strong negative correlations between inverse hypox classes are not expected to be seen in DO mice, which have an

intact pituitary gland and presumably maintain sex-specific plasma GH patterns are. Genetic variants that have accumulated in individual DO mice, an outbred population derived from 8 inbred founder mouse stains (Svenson et al. 2007, Churchill et al. 2012), must therefore alter a subset of regulatory regions that are responsible for the four hypoxresponse phenotypes. This hypothesis is supported by our finding that genetic variants are associated with gene expression changes in select correlated lncRNA-protein coding gene module pairs. In particular, we used the list of single nucleotide polymorphisms (SNPs) that are significantly associated with liver gene expression (expression quantitative trait loci, eQTLs) identified in Chapter 3. An example is the darkgrey cluster in male liver, a female class 2-enriched gene module, which negatively correlates with ncRNA inter chr8 7423, a male class 1 lncRNA (Table 4-2). eQTLs identified for both this lncRNA and for 7 of its 11 gene cluster members point to the genotype of the PWK/PhJ strain as being associated with the largest changes in expression of the proteincoding genes (Hao2, Cyp2a22, Sult1e1, Sybu, Tm4sf4, Setbp1 and Dcbld1; Table S4-4A). Further, eQTLs for 3 of the other 4 gene cluster members identified the PWK/PhJ strain as the strain whose genetic variants make the second to fourth largest contribution to the observed changes in each of the 3 protein-coding gene expression (Cyp2b9, Cyp2a4 and Cep112). Many of these eQTLs map to different chromosomes, consistent with this male-specific lncRNA being involved in *trans* regulation.

In conclusion, we expanded our earlier characterization of liver-expressed lcnRNAs by using gene expression data obtained under 30 distinct biological conditions, and identified many lncRNAs showing strong sex-biased gene expression. A subset of these sex-specific lncRNAs is dynamically regulated during postnatal liver development in male mice, and an overlapping subset shows significant oscillating gene expression patterns over a 24-hour period. We further identified sex-specific protein-coding gene clusters in male, and separately, female liver using a mouse population with rich genetic diversity (DO mouse), many of which showed unexpected enrichment for one of the four hypox-response classes. Sex-specific lncRNAs that are significantly correlated in expression with at least one protein-coding gene cluster were identified. Strikingly, a subset of sex-specific lncRNAs showed a significant negative correlation with proteincoding gene modules that were enriched for genes of the opposite sex-bias and inverse hypox classes, suggesting a negative regulatory role for these lncRNAs. These findings elucidate the sex-bias liver transcriptome, and illustrate the utility of gene modules to reveal broad characteristics of sex-bias gene regulation and generate hypothesis for regulatory roles for sex-bias lncRNAs.

#### **4.F Acknowledgements**

I thank Christine Nykyforchyn of the Waxman lab for her analysis of GH-responsive lncRNAs using the ribosomal-depleted RNA-Seq samples and her initial analysis of the gene expression correlations in DO mouse liver.

## Table 4-1. Enrichments of gene modules discovered in male, and separately, in female liver for sex-specific and hypox-responsive genes.

#Cluster columns indicate the number of cluster that are enriched for each respective gene set. The #Gene columns indicate the number of genes belonging to each respective gene set in these enriched cluster. The common genes row indicates the number of genes that overlap in male and female liver in each column.

Samples	Enri for M spec Gen	ched Male- ific es	Enrie for Fem spec Gene	ched ale- ific es	Enric for M Class Gene	ched Iale S I es	Enri for M Clas Gene	ched Male s II es	Enrie for Fem Clas Gene	ched ale s I es	Enriched for Female Class II Genes			
	#Cluster	#Gene	#Cluster	#Gene	#Cluster	#Gene	#Cluster	#Gene	#Cluster	#Gene	#Cluster	#Gene		
Male	13	240	10	10 379		85	2	12	3	21	6	89		
Female	12	201	9	339	6	68	3	11	5	23	3	82		
Common genes	143		265		56		6		7		68			

Liver lncRNA lncRN Pears Signifi Module Module Hypox #Ge cance classes sets 's nes А on correl of hypox enrich hypox in classes ation correla classes ment eac tion for coeffi h cient (FDR) each mo module dule ncRNA Male I -0.59 1.05E-Female 2.47E-13 Female darkora inter ch 09 nge Π 03 r7 6509 Male I Female 3.17E-57 Female ncRNA -0.45 3.82Egreen inter ch 06 05 Π r1 630 Female ncRNA Male I -0.42 1.34Edarkora Female 2.47E-13 inter ch 04 nge Π 03 r10\_931 3 Female ncRNA Male I -0.47 1.01Egreen Female 3.17E-57 inter ch 05 Π 06 r6 5316 ncRNA Male Male I -0.45 1.08Edarkgre Female 8.01E-11 inter ch 04 Π 04 V Female Female 5.70E-3.17Er8 7423 -0.44 57 green 05 Π 06 ncRNA Male 3.35E-Female 1.32E-Female -0.60grey60 18 inter ch 10 06 Π Ι r10 900 0 ncRNA 6.42E-Female Male -0.403.14Esteelblu Female 11 inter ch 04 03 Π e Ι r16 133 49 Female ncRNA Femal -0.53 1.73Ered Male II 2.17E-53 inter ch e I 07 03 r8 6946 Female ncRNA Femal -0.491.93E-06 inter ch e I r13\_114 37 2.90E-Female ncRNA Femal -0.41 04 inter ch e I

Table S4-2. List of sex-specific lncRNAs that are negatively correlated (FDR < 0.001) with protein-coding genes enriched (P < 0.01) for inverse hypox classes

	r16_134 28							
Male	ncRNA_ inter_ch r1_307	Femal e II	-0.40	8.55E- 04	lightcya n	Male I	1.91E- 04	23
Female	ncRNA_ inter_ch	Femal e II	-0.47	1.01E- 05	skyblue	Male I	1.68E- 03	12
Female	r10_922 2		-0.39	6.14E- 04	purple	Male I	8.23E- 03	25
Male	ncRNA_ inter_ch r16_135 10	Femal e II	-0.44	1.44E- 04	pink	Male I	3.57E- 05	40
Male	ncRNA_ inter_ch	Femal e II	-0.51	4.04E- 06				
Male	r2_1430		-0.48	1.18E- 05	lightcya n	Male I	1.91E- 04	23
Male			-0.43	1.82E- 04	sienna3	Male I	4.98E- 04	7
Male			-0.41	4.53E- 04	brown	Male I	9.58E- 11	64
Male	ncRNA_ inter_ch	Femal e II	-0.45	1.08E- 04	lightcya n	Male I	1.91E- 04	23
Male	r2_1451		-0.42	3.18E- 04	pink	Male I	3.57E- 05	40
Female	ncRNA_ inter_ch	Femal e II	-0.58	2.63E- 09	purple	Male I	8.23E- 03	25
Female	r3_2937		-0.54	1.27E- 07	greenye llow	Male I	1.96E- 06	22
Female			-0.41	2.62E- 04	skyblue	Male I	1.68E- 03	12

### Table S4-1 to Table S4-7

Supplementary tables for Chapter 4 are upon request (djw@bu.edu)

#### Figure 4-1. Liver-expressed lncRNAs.

(A) PCA of the 186 RNA-Seq samples used to discover liver-expressed lncRNAs. The first principal component corresponds to the age of the mice from young to old (right to left). (B) Examples of novel intergenic lncRNAs discovered in this study: ncRNA\_inter\_chr10\_9193, ncRNA\_inter\_chr10\_9194 and ncRNA\_inter\_chr10\_9195. Gene structures of lncRNAs are shown in green. Liver gene expression in select samples is shown as wig files. F, forward strand, R, reverse strand. (C) An antisense lncRNA, ncRNA\_as\_chr14\_12073, that is transcribed on the forward strand and overlaps *Gjb2*, which is transcribed on the reverse strand. (D) An intragenic lncRNA, ncRNA\_intra\_chr2\_1648, that is transcribed on the reverse strand, overlaps the gene *Gpr176* as well as another annotated lncRNA (1700054M17Rik).



## Figure 4-2. Sex-specific lncRNAs showing significant regulation during postnatal liver developmental and during a 24-hour light-dark cycle periods in male mouse liver.

(A) Heatmap of 144 sex-specific lncRNAs (male/female |fold-change| > 4 at FDR < 0.05in any of the 8 datasets assessed) that show a significant change in expression (|foldchange | > 2 and FDR < 0.05) at any time point from birth to maturity (Table S4-2A). The first 8 columns of the heatmap correspond to the 8 datasets to used to identify sexspecific lncRNAs; blue color indicates male-specific lncRNA expression, while red color indicates female-specific lncRNA expression. The last 10 columns track the change in expression of each lncRNA at a particular age, as compared to birth (day 0) in male liver. Blue denotes up regulation of gene expression, as compared to day 0, and red denotes down regulation. Set I designates male-specific genes that are up regulated after birth, set II are female-specific genes that are down regulated after birth, set III are male-specific genes that are down regulated after birth while maintaining their male-specific expression at adulthood, set IV are similar to set III genes, except they are up regulated around 30 days of age, and set V are female-specific genes that are up regulated by 25 days of age. See Table S4-2C for listings of lncRNAs in set I to V. (B) Heatmap of 52 sex-specific lncRNAs (male/female |fold-change| > 4 at FDR < 0.05 in any of the 8 datasets evaluated) that showed a significant change in expression (|fold-change| > 2 and FDR < 0.05) at one or more time points from ZT4-ZT22, as compared to ZT2 (Table S4-2B). The first 8 columns in the heatmap correspond to the 8 datasets used to identify sexspecific lncRNAs; blue color indicates male-specific gene expression, and red color indicates female-specific expression. The last 5 columns show the fold-change of expression at a particular ZT time, as compared to time ZT2. Blue indicates activation after ZT2; red indicates repression after ZT2. Set I and III are sex-specific lncRNAs that were up regulated at any time point after ZT2, and set II and IV are sex-specific lncRNAs that were repressed at any time point after ZT2. See Table S4-2D for listings of lncRNAs in set I to IV. (C) 24 h oscillating gene expression profiles are shown for four sexspecific lncRNAs, where the top figures show up regulation by ZT10, and the bottom figures show down regulation during the first 12 h.



### Figure 4-3. Sex-specific gene clusters (gene modules) discovered in DO male mouse liver.

(Top) Hierarchical clustering showing the relationship between all clusters, as defined by the correlation of their first principal component. Clusters are named by colors, as indicated. (Bottom) The first row indicates the number of genes in each cluster, and the second row indicates the number of distinct TADs represented in each cluster. The next three rows show the percentage of genes in each cluster that are either strongly sexspecific (FC>4; row 3), or that are male-specific (row 4) or female-specific genes (row 5). The next four rows indicate the percentage of genes distributed in each of the four hypox-response gene classes: male-specific class 1 and class 2, and female-specific class 1 and class 2; these percentages were calculated with respect to the subset of genes in each cluster that show a significant response to hypox, consequently, the sum of percentages for the last four rows in each cluster should equal to 100%. However, for rows 3-9, the percentage values are only listed for clusters that show a significant overlap with relevant gene sets in each row (p-value < 0.05, Hypergeometric test). Ovals indicate adjacent branches whose nodes are enriched for the same hypox-response gene class. Clusters are named by *colors*, as indicated; in several cases the cluster name (i.e., color) assigned to a cluster, discovered in DO male mouse liver is the same name as that assigned to one of the clusters discovered in DO female mouse liver and shown in Fig. 4-4; however, there is no relationship between such clusters. Thus, male DO mouse liver cluster Cyan, with 71 genes, displayed here, is not related to female DO mouse liver cluster Cyan, with 20 genes, shown in Fig. 4-4. Cluster A consists of protein-coding gene modules enriched for female-specific genes while cluster B consists of protein-coding gene modules enriched for male-specific genes.



Fig. 4-3

#### Figure 4-4. Sex-specific gene clusters discovered in DO female mouse liver.

Characteristics of the gene clusters shown are displayed as described in the legend to Fig. 4-3. Cluster A consists of protein-coding gene modules enriched for female-specific genes while cluster B1 and B2 consist of protein-coding gene modules enriched for male-specific genes.





# Figure 4-5. Overlap between clusters discovered in male *vs.* female liver for clusters that are significantly enriched for sex-specific genes and/or hypox-response gene classes.

The number shown next to each cluster name (designated as a *color*; see legend to Fig. 4-3) represents the number of genes in that cluster. Clusters discovered in DO male liver are shown along the Y-axis; those discovered in DO female liver are shown along the Xaxis. The matrix shows the number of genes common to the intersecting gene clusters discovered in the DO livers from each sex. Shades of red (bar at *right*) represent the percentage of overlap with respect to the cluster discovered in male liver. Circles denote cluster pairs with strong gene overlap, defined as >50% of gene members in a specific cluster in DO male livers (row) being present in a specific cluster in DO female livers (column).



																												%	0	verlap
	areen: 54	0	0	2	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1(	40	2	0	0	N/	' (	luster
s	blue: 84	0	0	0	2	0	0	0	0	1	0	0	0	0	0	1	0	0	2	0	2	0	1	62	2	1	2	l İl	n I	Male
t	cyan: 70	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	1	51	4	0	1			
ne	darkgreen: 12	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	0	2	2	0	1			
Ę	darkorange: 11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	3	1	4	0			
Ľ.	steelblue: 10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0		-	60
Ē	midnightblue: 23	2	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	3	1	0	0			
ine	darkgrey: 11	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	4	1	0	0	0			
Ľ,	turquoise: 137	0	0	0	1	0	0	0	3	0	2	0	0	0	2	3	1	7	7	9	6	9	27	12	1	0	0			
in in	yellow: 59	0	0	0	0	0	0	1	2	0	0	2	1	1	0	1	0	0	1	2	1	26	1	3	0	0	0			
Si	magenta: 34	0	7	1	0	0	0	1	2	0	0	2	0	0	0	0	1	1	1	0	0	0	1	12	0	0	0			
∑	skyblue: 10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	5	0	0	0	0		-	40
٩L	grey60: 21	1	1	6	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0			
÷	saddlebrown: 10	0	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	1	0	0	0			
Š	tan: 30	0	0	0	2	4	2	0	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0	5	3	0	0			
<u>v</u>	salmon: 29	2	0	3	1	2	1	0	4	0	0	0	1	1	0	0	1	2	0	0	0	0	0	9	0	0	0			
 	darkturquoise: 12	0	0	0	2	5	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
ale	violet: 9	0	0	0	0	2	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	1	0	0	0		-	20
Σ	brown: 62	3	1	1	2	0	3	0	21	0	1	1	1	0	1	0	0	0	0	0	0	2	0	11	0	0	1			
⊒.	lightcyan: 23	0	1	0	3	3	0	0	2	0	1	0	0	0	2	0	0	0	0	0	0	2	0	3	0	0	0			
ers	pink: 40	2	2	0	2	1	2	0	13	0	2	0	3	1	0	0	0	0	0	0	0	1	0	5	0	0	1			
Ĩ	sienna3: 7	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0			
ы	lightgreen: 21	1	2	0	0	1	1	0	1	0	0	0	0	0	7	0	0	0	0	0	0	0	0	3	0	0	0			
	lightsteelblue1: 5	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0			0
	skyblue3: 6	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	ļ		
		in: 20	n: 21	sy: 14	w: 22	le: 25	le: 12	le3: 7	n: 64	le4: 5	et: 11	en: 8	en: 18	lite: 5	d: 14	le1: 7	le: 11	le: 15	0: 18	le: 14	le: 13	id: 53	in: 57	: 271	w: 18	ia: 10	n1: 8			
		cya	salmo	darkgre	greenyello	purpl	skyblu	mediumpurpl	brow	bisdu	viole	yellowgre	lightgree	floralwh	darkre	lightsteelblu	steelblu	royalblu	grey6	orang	darkorang	Le	gree	black	lightyello	darkmagent	plur			

Clusters in Female Liver with Any Significant Enrichments

### Figure 4-6. Schematic diagram of the eight proposed scenarios for negative regulatory role of sex-specific lncRNAs.

(Left) Gene regulatory mechanisms in male liver: the male, pulsatile plasma GH pattern activates male class 1 lncRNAs in male liver, which are proposed to repress female-class 2 protein-coding genes in male liver (*first row*). Similarly, male-class 2 lncRNAs, which are constitutively expressed in male liver, are proposed to repress female class 1 proteincoding genes in male liver (second row). Female class 1 lncRNAs are constitutively off in male liver, and therefore cannot repress male-class 2 protein-coding genes (third row). Female class 2 lncRNAs, whose expression is repressed by the male plasma GH pattern in male liver, therefore cannot repress male class 1 protein-coding genes (fourth row). (Bottom) Gene regulatory mechanisms in female liver: in the absence of pulsatile GH, male class 1 lncRNAs are constitutively off in female liver, and thus cannot repress female-class 2 protein-coding genes (top row). Male-class 2 lncRNAs are repressed by the female plasma GH pattern, and cannot therefore repress female class 1 protein-coding genes (second row). The female, persistent plasma GH pattern activates female class 1 lncRNAs, which are proposed to repress male-class 2 protein-coding genes (third row). Female-class 2 lncRNAs are constitutively on in female liver, and can therefore repress male class 1 protein-coding genes (fourth row). The inverse regulatory relationship between class 1 sex-specific lncRNAs and class 2 sex-specific protein coding genes shown here is consistent with the effects of hypox on both classes of sex-specific genes. For example, hypox of male mice, which abolishes the GH pulse regulatory circuit diagrammed on the left, down regulates class 1 male lncRNAs and up regulates (derepresses) female class 2 protein coding genes in male liver. That de-repression is proposed to involve the loss of the male class 1 lncRNAs required for female class 2 protein-coding gene repression. Hypox of female mice, which abolishes the persistent GH regulatory circuit diagrammed on the right, down regulates class 1 female lncRNAs and up regulates (de-represses) male class 2 protein coding genes in female liver.


# Figure 4-7. Distribution of hypox-response gene classes for intergenic, sex-specific lncRNAs whose expression in DO mouse liver is eitherpositively or negatively correlated with protein-coding gene modules that are enriched for one the four hypox-response classes.

(A) Shown is the distribution of hypox-response classes for the 142 multi-exonic, intergenic sex-specific lncRNAs that are varying (MAD > 0) in male liver (first bar, marked All), 79 sex-specific lncRNAs whose expression is positively correlated with any protein-coding gene module enriched for hypox-responsive genes in male liver (*left*) and 63 sex-specific lncRNAs that are negatively correlated with any protein-coding gene module enriched for hypox-responsive genes in male liver (right). Gene modules in the left and right panels are both ordered by their enrichment for male class 1 protein-coding genes from left to right (low FDR to high FDR). More male class 1 lncRNAs are seen to be positively correlated with protein-coding genes that are strongly enriched for male class 1 genes, whereas more female class 2 lncRNAs are negatively correlated with protein coding gene clusters enriched for male class 1 genes. (B) Shown is the distribution of hypox-response classes for the 143 multi-exonic, intergenic sex-specific lncRNAs that are varying (MAD > 0) in female liver (first bar, marked All), 84 sexspecific lncRNAs that are positively correlated with any protein-coding gene modules enriched for hypox-responsive genes in female liver (left), and 59 sex-specific lncRNAs that are negatively correlated with any protein-coding gene modules enriched for hypoxresponsive genes in female liver (right). Gene modules in the *left* and *right* panels are both ordered by their enrichment for female class 1 protein-coding genes from left to right (low FDR to high FDR). The number in parenthesis after each gene cluster name indicates the total number of lncRNAs whose hypox class distributions are shown in each bar graph. More female class 1 lncRNAs positively correlate with protein-coding genes that are strongly enriched for female class 1 genes, and more, albeit to a lesser extent, male class 2 lncRNAs negatively correlate with protein coding gene clusters enriched for female class 1 genes.



## Figure S4-1. Four hypox-response classes.

Genes in male hypox class 1 and female hypox class 1 are activated by the GH secretion pattern of the sex where they are more highly expressed, whereas, male hypox class 2 and female hypox class 2 genes are repressed by the GH secretion pattern of the sex where they are less highly expressed. Consequently, following hypox, class 1 genes are down regulated and class 2 genes are de-repressed (up regulated).











**Figure S4-3. A Heatmap of 684 sex-specific lncRNAs in different RNA-Seq datasets.** The first 8 columns in the heatmap correspond to the 8 datasets used to identify sexspecific lncRNAs; blue indicates male-specific gene expression, and red indicates female-specific gene expression. The last 10 columns track the changes of gene expression in male liver at the ages specified at top, as compared to birth (day 0). Blue denotes up regulation of gene expression, compared to day 0, and red denotes down regulation of gene expression at the specified age, as compared to day 0.



Fig. S4-3

### Figure S4-4. Overlap between clusters discovered in male vs. female liver.

The number of genes in each matrix cell represents the number of genes in the cluster discovered in male DO mouse liver (y-axis) that are present in the indicated clusters discovered in female DO mouse liver (x-axis). The last cluster in female: grey, is not considered a cluster; it contains 1 gene that was not able to be clustered in female liver. The color represents the percentage of overlap with respect to the cluster discovered in male liver. The number written along each axis after each cluster name is the number of genes in each cluster. Circles denote strong overlap, where >50% of gene members in a specific cluster in male (row) are in a specific cluster in female (column). Also, see Fig. 4-5.





## Figure S4-5. Examples of gene modules that are negatively correlated with sexspecific lncRNAs (Table 4-2).

Nodes denote genes, and edges denote the correlation between genes. The darkness of the edges indicates the correlation strength between genes. The size of each node represents the cumulative gene correlations for a gene, where the larger the node is, the more high correlations involve that node. Graphs were drawn using Cytoscape (Smoot et al. 2011). Pairwise Pearson correlation for select edges are shown in red. Pairwise correlations between gene members with the lncRNA with which the respective module is correlated are tabulated for select gene modules.

## Fig. S4-5



Pairwise Pears	on correlation
with ncRNA_in	ter_chr8_7423
Cyp2b9	_0 52 → Strongest

Cyp2b9	-0.52 -> Strongest neg.
Cyp2b13	-0.41 corr. genome-
Hao2	-0.39
Sybu	-0.25
Cyp2a22	-0.24
Cyp2a4	-0.21
Sult1e1	-0.19
Tm4sf4	-0.17
Setbp1	-0.16
Cep112	-0.14
Dcbld1	-0.07

Pairwise Pearson cor. w/ Pairwise Pearson cor. w/

Skyblue cluster in female DO Cor. w/ncRNA\_inter\_chr3\_2937: -0.41 Cor. w/ncRNA\_inter\_chr10\_9222: -0.47

Cor. w/ncRNA_inter_chr10_9222: -0.47	ncRNA_inter_c Max genome-w cor. = - <mark>0.52</mark> (C8	hr3_2937 vide neg. a)	ncRNA_inter Max genome cor. = -0.44 (9	_chr10_9222 -wide neg. C8b)
Ces4a 0.4	Egfr	-0.48	Egfr	-0.43
Serpinag	Hsd3b5	-0.33	Cmah	-0.41
Gpc1 Hsd3b5 0.4 Slc22	a28 Serpina9	-0.33	Lrrc16a	-0.38
	Farp1	-0.28	Serpina9	-0.36
0.14	Lrrc16a	-0.26	Hsd3b5	-0.32
Egfr	Cmah	-0.23	Farp1	-0.31
LITICIDA	Ces4a	-0.20	Slc17a2	-0.25
SI-17-2	Treh	-0.17	Slc22a28	-0.25
SICITAZ	Slc17a2	-0.17	Ces4a	-0.12
Treh 0.15	Gpc1	-0.13	Wisp3	-0.09
	Wisp3	-0.08	Treh	-0.08
Wisp3	Slc22a28	-0.04	Gpc1	-0.07



#### LIST OF JOURNAL ABBREVIATIONS

AAPS J The AAPS journal The American journal of physiology Am J Physiol American journal of physiology. Endocrinology and Am J Physiol Endocrinol Metab metabolism American Physiological Society Ann N Y Acad Sci Annals of the New York Academy of Sciences Annual review of biochemistry Annu Rev Biochem Arch Med Res Archives of Medical Research Biochemistry **Biochemistry Bioinformatics Bioinformatics** Biology direct **Biol Direct Biol Sex Differ Biology of Sex Differences** bioRxiv bioRxiv **BMC** Bioinformatics **BMC** bioinformatics **BMC** Genomics BMC genomics Breast Cancer Res Breast cancer research : BCR Cancer Lett Cancer letters Cancer Res Cancer Research Carcinogenesis Carcinogenesis Cell Cell Cell Mol Life Sci Cellular and molecular life sciences Cell reports Cell Rep Chronobiology international Chronobiol Int Comprehensive Physiology **Comprehensive Physiology** Curr Opin Cardiol Current Opinion in Cardiology Dev Cell Developmental cell Development Development Drug Discovery Today Drug Discovery Today Drug Metabolism Reviews Drug Metabolism Reviews Elife eLife Endocr Rev Endocrine reviews Endocrinology Endocrinology **Environ Health Perspect Environmental Health Perspectives** Expert Opin Drug Metab Toxicol Expert Opinion on Drug Metabolism & Toxicology Expert Rev Mol Med Expert Reviews in Molecular Medicine FASEB J Federation of American Societies for Experimental **Biology Journal** Food Chem Toxicol Food and Chemical Toxicology Genes, Genomes, Genetics G3 Gastroenterology Gastroenterology Genes Dev Genes & Development Genetics Genetics Genome Biol Genome Biology

Genome Biol Evol	Genome Biology and Evolution
Genome Res	Genome research
Genomics	Genomics
Growth Horm IGF Res	Growth hormone and IGF research
Hepatology	Hepatology
Hum Mol Genet	Human Molecular Genetics
Ilar i	ILAR journal
Int J Biochem Cell Biol	The International Journal of Biochemistry & Cell
	Biology
Int J Cancer	International Journal of Cancer
Int J Clin Pharmacol Ther	International Journal of Clinical Pharmacology and
	Therapeutics
J Appl Physiol (1985)	Journal of Applied Physiology
J Biol Chem	The Journal of Biological Chemistry
J Clin Endocrinol Metab	The Journal of Clinical Endocrinology and
	Metabolism
J Clin Pharm Ther	Journal of Clinical Pharmacy and Therapeutics
J Endocrinol	The Journal of Endocrinology
J Endocrinol Invest	Journal of Endocrinological Investigation
J Neurosci	The Journal of Neuroscience
J Pharmacol Exp Ther	The Journal of Pharmacology and Experimental
F F	Therapeutics
J Stat Softw	Journal of Statistical Software
J Womens Health Gend Based Med	Journal of Women's Health & Gender-based
	Medicine
JAK-STAT	Jakstat
J Mol Med	Journal of Molecular Medicine
Machine Learning	Machine Learning
Mamm Genome	Mammalian Genome
Mediators Inflamm	Mediators of Inflammation
Mol Cell	Molecular Cell
Mol Cell Biol	Molecular and Cellular Biology
Mol Endocrinol	Molecular Endocrinology
Mol Pharmacol	Molecular Pharmacology
Mutat Res	Mutation Research
Nat Biotechnol	Nature Biotechnology
Nat Genet	Nature Genetics
Nat Methods	Nature Methods
Nat Protoc	Nature Protocols
Nat Rev Endocrinol	Nature Reviews Endocrinology
Nat Struct Mol Biol	Nature Structural & Molecular Biology
Nature	Nature
Nucleic Acids Res	Nucleic Acids Research
Nutr Metab	Nutrition and Metabolism

Oncogene Oncogene Pharmacol Ther Pharmacology & Therapeutics Pharmacological Reviews Pharmacol Rev Physiological Genomics **Physiol Genomics** Plant Physiol Plant Physiology PLoS Genet PLoS Genetics PLoS One PloS One Proceedings of the National Academy of Sciences Proc Natl Acad Sci U S A of the United States of America **RNA** Rna Scientific reports Sci Rep Sci Signal Science signaling Science Science Stat Appl Genet Mol Biol Statistical Applications in Genetics and Molecular Biology Toxicology letters **Toxicol Lett** Toxicological sciences Toxicol Sci **Toxicol Mech Methods** Toxicology Mechanisms and Methods Trends in Genetics Trends Genet Trends in Biochemical Sciences Trends Biochem Sci World J Gastroenterol World Journal of Gastroenterology

#### **BIBLIOGRAPHY**

Adams, J. M., V. Otero-Corchon, G. L. Hammond, J. D. Veldhuis, N. Qi and M. J. Low (2015). "Somatostatin Is Essential for the Sexual Dimorphism of GH Secretion, Corticosteroid-Binding Globulin Production, and Corticosterone Levels in Mice." <u>Endocrinology</u> **156**(3): 1052-1065.

Alam, T., Y. A. Medvedeva, H. Jia, J. B. Brown, L. Lipovich and V. B. Bajic (2014). "Promoter analysis reveals globally differential regulation of human long non-coding RNA and protein-coding genes." <u>PLoS One</u> **9**(10): e109443.

Alizadeh, A. A., M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C.
Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J.
Hudson, Jr., L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner,
D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C.
Byrd, D. Botstein, P. O. Brown and L. M. Staudt (2000). "Distinct types of diffuse large
B-cell lymphoma identified by gene expression profiling." Nature 403(6769): 503-511.

Amacher, D. E. (2010). "The effects of cytochrome P450 induction by xenobiotics on endobiotic metabolism in pre-clinical safety studies." <u>Toxicology mechanisms and</u> <u>methods</u> **20**(4): 159-166.

Amaral, P. P., M. B. Clark, D. K. Gascoigne, M. E. Dinger and J. S. Mattick (2011). "IncRNAdb: a reference database for long noncoding RNAs." <u>Nucleic Acids Res</u> **39**(Database issue): D146-151.

Amit-Avraham, I., G. Pozner, S. Eshar, Y. Fastman, N. Kolevzon, E. Yavin and R. Dzikowski (2015). "Antisense long noncoding RNAs regulate var gene activation in the malaria parasite Plasmodium falciparum." <u>Proc Natl Acad Sci U S A</u> **112**(9): E982-991.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." <u>Genome Biol</u> **11**(10): R106.

Anders, S., P. T. Pyl and W. Huber (2015). "HTSeq-a Python framework to work with high-throughput sequencing data." <u>Bioinformatics</u> **31**(2): 166-169.

Anthony, M. and M. J. Berg (2002). "Biologic and molecular mechanisms for sex differences in pharmacokinetics, pharmacodynamics, and pharmacogenetics: Part I." J Womens Health Gend Based Med **11**(7): 601-615.

Anwar, S. L., T. Krech, B. Hasemeier, E. Schipper, N. Schweitzer, A. Vogel, H. Kreipe and U. Lehmann (2012). "Loss of imprinting and allelic switching at the DLK1-MEG3 locus in human hepatocellular carcinoma." <u>PLoS One</u> 7(11): e49462.

Ayonrinde, O. T., J. K. Olynyk, L. J. Beilin, T. A. Mori, C. E. Pennell, N. de Klerk, W. H. Oddy, P. Shipman and L. A. Adams (2011). "Gender-specific differences in adipose distribution and adipocytokines influence adolescent nonalcoholic fatty liver disease." <u>Hepatology</u> **53**(3): 800-809.

Baik, M., J. H. Yu and L. Hennighausen (2011). "Growth hormone-STAT5 regulation of growth, hepatocellular carcinoma, and liver metabolism." <u>Ann N Y Acad Sci</u> **1229**: 29-37.

Balwierz, P. J., P. Carninci, C. O. Daub, J. Kawai, Y. Hayashizaki, W. Van Belle, C. Beisel and E. van Nimwegen (2009). "Methods for analyzing deep sequencing expression data: constructing the human and mouse promoterome with deepCAGE data." <u>Genome Biol</u> **10**(7): R79.

Banerjee, M., D. Robbins and T. Chen (2014). "Targeting xenobiotic receptors PXR and CAR in human diseases." <u>Drug discovery today</u>.

Bavia, L., I. A. de Castro and L. Isaac (2015). "C57BL/6 and A/J Mice Have Different Inflammatory Response and Liver Lipid Profile in Experimental Alcoholic Liver Disease." <u>Mediators Inflamm</u> **2015**: 491641.

Bhattacharjee, A., W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker and M. Meyerson (2001). "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses." <u>Proc Natl Acad Sci U S A</u> **98**(24): 13790-13795.

Bittner, V. (2006). "Perspectives on dyslipidemia and coronary heart disease in women: an update." <u>Curr Opin Cardiol</u> **21**(6): 602-607.

Boergesen, M., T. A. Pedersen, B. Gross, S. J. van Heeringen, D. Hagenbeek, C. Bindesboll, S. Caron, F. Lalloyer, K. R. Steffensen, H. I. Nebb, J. A. Gustafsson, H. G. Stunnenberg, B. Staels and S. Mandrup (2012). "Genome-wide profiling of liver X receptor, retinoid X receptor, and peroxisome proliferator-activated receptor alpha in mouse liver reveals extensive sharing of binding sites." <u>Mol Cell Biol</u> **32**(4): 852-867.

Bogue, M. A., G. A. Churchill and E. J. Chesler (2015). "Collaborative Cross and Diversity Outbred data resources in the Mouse Phenome Database." <u>Mamm Genome</u> **26**(9-10): 511-520.

Braconi, C., T. Kogure, N. Valeri, N. Huang, G. Nuovo, S. Costinean, M. Negrini, E. Miotto, C. M. Croce and T. Patel (2011). "microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer." <u>Oncogene</u> **30**(47): 4750-4756.

Brooks, A. J., W. Dai, M. L. O'Mara, D. Abankwa, Y. Chhabra, R. A. Pelekanos, O. Gardon, K. A. Tunny, K. M. Blucher, C. J. Morton, M. W. Parker, E. Sierecki, Y. Gambin, G. A. Gomez, K. Alexandrov, I. A. Wilson, M. Doxastakis, A. E. Mark and M. J. Waters (2014). "Mechanism of activation of protein kinase JAK2 by the growth hormone receptor." <u>Science</u> **344**(6185): 1249783.

Brooks, A. J. and M. J. Waters (2010). "The growth hormone receptor: mechanism of activation and clinical implications." <u>Nat Rev Endocrinol</u> **6**(9): 515-525.

Cabili, M. N., C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev and J. L. Rinn (2011). "Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses." <u>Genes Dev</u> **25**(18): 1915-1927.

Chesler, E. J., D. M. Gatti, A. P. Morgan, M. Strobel, L. Trepanier, D. Oberbeck, S. McWeeney, R. Hitzemann, M. Ferris, R. McMullan, A. Clayshultle, T. A. Bell, F. P. Manuel de Villena and G. A. Churchill (2016). "Diversity Outbred Mice at 21: Maintaining Allelic Variation in the Face of Selection." G3 (Bethesda) **6**(12): 3893-3902.

Cheung, C., A. M. Yu, C. S. Chen, K. W. Krausz, L. G. Byrd, L. Feigenbaum, R. J. Edwards, D. J. Waxman and F. J. Gonzalez (2006). "Growth hormone determines sexual dimorphism of hepatic cytochrome P450 3A4 expression in transgenic mice." J Pharmacol Exp Ther **316**(3): 1328-1334.

Chia, D. J., M. Ono, J. Woelfle, M. Schlesinger-Massart, H. Jiang and P. Rotwein (2006). "Characterization of distinct Stat5b binding sites that mediate growth hormone-stimulated IGF-I gene transcription." J Biol Chem **281**(6): 3190-3197.

Chick, J. M., S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson, G. A. Churchill and S. P. Gygi (2016). "Defining the consequences of genetic variation on a proteome-wide scale." <u>Nature</u> **534**(7608): 500-505.

Choi, H. K. and D. J. Waxman (2000). "Plasma growth hormone pulse activation of hepatic JAK-STAT5 signaling: developmental regulation and role in male-specific liver gene expression." <u>Endocrinology</u> **141**(9): 3245-3255.

Church, R. J., D. M. Gatti, T. J. Urban, N. Long, X. Yang, Q. Shi, J. S. Eaddy, M. Mosedale, S. Ballard, G. A. Churchill, V. Navarro, P. B. Watkins, D. W. Threadgill and A. H. Harrill (2015). "Sensitivity to hepatotoxicity due to epigallocatechin gallate is affected by genetic background in diversity outbred mice." Food Chem Toxicol **76**: 19-26.

Churchill, G. A., D. M. Gatti, S. C. Munger and K. L. Svenson (2012). "The Diversity Outbred mouse population." <u>Mamm Genome</u> **23**(9-10): 713-718.

Clodfelter, K. H., M. G. Holloway, P. Hodor, S. H. Park, W. J. Ray and D. J. Waxman (2006). "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." <u>Mol Endocrinol</u> **20**(6): 1333-1351.

Clodfelter, K. H., G. D. Miles, V. Wauthier, M. G. Holloway, X. Zhang, P. Hodor, W. J. Ray and D. J. Waxman (2007). "Role of STAT5a in regulation of sex-specific gene expression in female but not male mouse liver revealed by microarray analysis." <u>Physiol Genomics</u> **31**(1): 63-74.

Conforto, T. L., G. F. t. Steinhardt and D. J. Waxman (2015). "Cross Talk Between GH-Regulated Transcription Factors HNF6 and CUX2 in Adult Mouse Liver." <u>Mol</u> <u>Endocrinol</u> **29**(9): 1286-1302.

Conforto, T. L., G. F. t. Steinhardt and D. J. Waxman (2015). "Crosstalk between Growth Hormone-regulated Transcription Factors HNF6 and CUX2 in Adult Mouse Liver." <u>Molecular endocrinology</u>: me20151028.

Conforto, T. L., G. F. t. Steinhardt and D. J. Waxman (2015). "Crosstalk between Growth Hormone-regulated Transcription Factors HNF6 and CUX2 in Adult Mouse Liver." <u>Mol Endocrinol</u>: me20151028.

Conforto, T. L. and D. J. Waxman (2012). "Sex-specific mouse liver gene expression: genome-wide analysis of developmental changes from pre-pubertal period to young adulthood." <u>Biol Sex Differ</u> **3**: 9.

Conforto, T. L., Y. Zhang, J. Sherman and D. J. Waxman (2012). "Impact of CUX2 on the female mouse liver transcriptome: activation of female-biased genes and repression of male-biased genes." <u>Mol Cell Biol</u> **32**(22): 4611-4627.

Connerney, J., D. Lau-Corona, A. Rampersaud and D. J. Waxman (2017). "Activation of Male Liver Chromatin Accessibility and STAT5-Dependent Gene Transcription by Plasma Growth Hormone Pulses." <u>Endocrinology</u> **158**(5): 1386-1405.

Consortium, C. C. (2012). "The genome architecture of the Collaborative Cross mouse genetic reference population." Genetics **190**(2): 389-401.

Cordoba-Chacon, J., M. D. Gahete, J. P. Castano, R. D. Kineman and R. M. Luque (2011). "Somatostatin and its receptors contribute in a tissue-specific manner to the sexdependent metabolic (fed/fasting) control of growth hormone axis in mice." <u>American</u> journal of physiology. Endocrinology and metabolism **300**(1): E46-54.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R.

Jaenisch (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." <u>Proceedings of the National Academy of Sciences of the United</u> <u>States of America</u> **107**(50): 21931-21936.

Dardente, H. and N. Cermakian (2007). "Molecular circadian rhythms in central and peripheral clocks in mammals." <u>Chronobiol Int</u> **24**(2): 195-213.

Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow and R. Guigo (2012). "The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression." <u>Genome Res</u> **22**(9): 1775-1789.

Derti, A., P. Garrett-Engele, K. D. Macisaac, R. C. Stevens, S. Sriram, R. Chen, C. A. Rohl, J. M. Johnson and T. Babak (2012). "A quantitative atlas of polyadenylation in five mammals." <u>Genome Res</u> **22**(6): 1173-1183.

Eden, S. (1979). "Age- and sex-related differences in episodic growth hormone secretion in the rat." <u>Endocrinology</u> **105**(2): 555-560.

Eising, E., N. Pelzer, L. S. Vijfhuizen, B. Vries, M. D. Ferrari, P. A. t Hoen, G. M. Terwindt and A. M. van den Maagdenberg (2017). "Identifying a gene expression signature of cluster headache in blood." <u>Sci Rep</u> 7: 40218.

Engreitz, J. M., A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath and M. Guttman (2013). "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome." <u>Science</u> **341**(6147): 1237973.

Ernst, J. and M. Kellis (2013). "Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types." <u>Genome Res</u> **23**(7): 1142-1154.

Everett, L. J., J. Le Lay, S. Lukovac, D. Bernstein, D. J. Steger, M. A. Lazar and K. H. Kaestner (2013). "Integrative genomic analysis of CREB defines a critical role for transcription factor networks in mediating the fed/fasted switch in liver." <u>BMC Genomics</u> **14**: 337.

Fang, B., L. J. Everett, J. Jager, E. Briggs, S. M. Armour, D. Feng, A. Roy, Z. Gerhart-Hines, Z. Sun and M. A. Lazar (2014). "Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo." <u>Cell</u> **159**(5): 1140-1152.

Faure, A. J., D. Schmidt, S. Watt, P. C. Schwalie, M. D. Wilson, H. Xu, R. G. Ramsay, D. T. Odom and P. Flicek (2012). "Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules." <u>Genome Res</u> **22**(11): 2163-2175.

Fengler, V. H., T. Macheiner, S. M. Kessler, B. Czepukojc, K. Gemperlein, R. Muller, A. K. Kiemer, C. Magnes, J. Haybaeck, C. Lackner and K. Sargsyan (2016). "Susceptibility of Different Mouse Wild Type Strains to Develop Diet-Induced NAFLD/AFLD-Associated Liver Disease." <u>PLoS One</u> **11**(5): e0155163.

Ficklin, S. P., F. Luo and F. A. Feltus (2010). "The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks." <u>Plant Physiol</u> **154**(1): 13-24.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. Sonnhammer, J. Tate and M. Punta (2014). "Pfam: the protein families database." <u>Nucleic Acids Res</u> **42**(Database issue): D222-230.

Finn, R. D., J. Clements and S. R. Eddy (2011). "HMMER web server: interactive sequence similarity searching." <u>Nucleic Acids Res</u> **39**(Web Server issue): W29-37.

Friedman, J. R. and K. H. Kaestner (2006). "The Foxa family of transcription factors in development and metabolism." <u>Cell Mol Life Sci</u> **63**(19-20): 2317-2328.

Furth, P. A., R. E. Nakles, S. Millman, E. S. Diaz-Cruz and M. C. Cabrera (2011). "Signal transducer and activator of transcription 5 as a key signaling pathway in normal mammary gland developmental biology and breast cancer." <u>Breast cancer research : BCR</u> **13**(5): 220.

Gatti, D., A. Maki, E. J. Chesler, R. Kirova, O. Kosyk, L. Lu, K. F. Manly, R. W. Williams, A. Perkins, M. A. Langston, D. W. Threadgill and I. Rusyn (2007). "Genomelevel analysis of genetic regulation of liver gene expression networks." <u>Hepatology</u> **46**(2): 548-557.

Gatti, D. M., P. Simecek, L. Somes, C. T. Jeffery, M. J. Vincent, K. Choi, X. Chen, G. A. Churchill and K. L. Svenson (2017). "The Effects of Sex and Diet on Physiology and Liver Gene Expression in Diversity Outbred Mice." <u>bioRxiv</u>.

Gatti, D. M., K. L. Svenson, A. Shabalin, L. Y. Wu, W. Valdar, P. Simecek, N. Goodwin, R. Cheng, D. Pomp, A. Palmer, E. J. Chesler, K. W. Broman and G. A. Churchill (2014). "Quantitative trait locus mapping methods for diversity outbred mice." <u>G3 (Bethesda)</u> 4(9): 1623-1633.

Gatti, D. M., N. Zhao, E. J. Chesler, B. U. Bradford, A. A. Shabalin, R. Yordanova, L. Lu and I. Rusyn (2010). "Sex-specific gene expression in the BXD mouse liver." <u>Physiol</u> <u>Genomics</u> **42**(3): 456-468.

Gayen, S., E. Maclary, E. Buttigieg, M. Hinten and S. Kalantry (2015). "A Primary Role for the Tsix lncRNA in Maintaining Random X-Chromosome Inactivation." <u>Cell Rep</u> **11**(8): 1251-1265.

Gebert, C. A., S. H. Park and D. J. Waxman (1999). "Termination of growth hormone pulse-induced STAT5b signaling." <u>Mol Endocrinol</u> **13**(1): 38-56.

Gebert, C. A., S. H. Park and D. J. Waxman (1999). "Termination of growth hormone pulse-induced STAT5b signaling." <u>Molecular endocrinology</u> **13**(1): 38-56.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis,
L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F.
Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.
Y. Yang and J. Zhang (2004). "Bioconductor: open software development for computational biology and bioinformatics." Genome Biol 5(10): R80.

Gong, C. and L. E. Maquat (2011). "lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements." <u>Nature</u> **470**(7333): 284-288.

Gonzalez, F. J. (2005). "Role of cytochromes P450 in chemical toxicity and oxidative stress: studies with CYP2E1." <u>Mutat Res</u> **569**(1-2): 101-110.

Gordillo, M., T. Evans and V. Gouon-Evans (2015). "Orchestrating liver development." <u>Development</u> **142**(12): 2094-2108.

Gu, T., D. M. Gatti, A. Srivastava, E. M. Snyder, N. Raghupathy, P. Simecek, K. L. Svenson, I. Dotu, J. H. Chuang, M. P. Keller, A. D. Attie, R. E. Braun and G. A. Churchill (2016). "Genetic Architectures of Quantitative Variation in RNA Editing Pathways." <u>Genetics</u> **202**(2): 787-798.

Gu, X. and J. E. Manautou (2012). "Molecular mechanisms underlying chemical liver injury." <u>Expert reviews in molecular medicine</u> 14: e4.

Guengerich, F. P. (2006). "Cytochrome P450s and other enzymes in drug metabolism and toxicity." <u>AAPS J</u> 8(1): E101-111.

Guil, S. and M. Esteller (2012). "Cis-acting noncoding RNAs: friends and foes." <u>Nature</u> structural & molecular biology **19**(11): 1068-1075.

Gur, R. C., B. I. Turetsky, M. Matsui, M. Yan, W. Bilker, P. Hughett and R. E. Gur (1999). "Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance." J Neurosci **19**(10): 4065-4072.

Guttman, M., I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn and E. S. Lander (2009). "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals." Nature **458**(7235): 223-227.

Guttman, M., J. Donaghey, B. W. Carey, M. Garber, J. K. Grenier, G. Munson, G. Young, A. B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J. L. Rinn, D. E. Root and E. S. Lander (2011). "lincRNAs act in the circuitry controlling pluripotency and differentiation." <u>Nature</u> **477**(7364): 295-300.

Guttman, M., M. Garber, J. Z. Levin, J. Donaghey, J. Robinson, X. Adiconis, L. Fan, M. J. Koziol, A. Gnirke, C. Nusbaum, J. L. Rinn, E. S. Lander and A. Regev (2010). "Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs." <u>Nat Biotechnol</u> **28**(5): 503-510.

Hacisuleyman, E., L. A. Goff, C. Trapnell, A. Williams, J. Henao-Mejia, L. Sun, P.
McClanahan, D. G. Hendrickson, M. Sauvageau, D. R. Kelley, M. Morse, J. Engreitz, E.
S. Lander, M. Guttman, H. F. Lodish, R. Flavell, A. Raj and J. L. Rinn (2014).
"Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre." <u>Nat Struct Mol Biol</u> 21(2): 198-206.

Halley, P., B. M. Kadakkuzha, M. A. Faghihi, M. Magistri, Z. Zeier, O. Khorkova, C. Coito, J. Hsiao, M. Lawrence and C. Wahlestedt (2014). "Regulation of the apolipoprotein gene cluster by a long noncoding RNA." <u>Cell Rep</u> **6**(1): 222-230.

He, Y., X. M. Meng, C. Huang, B. M. Wu, L. Zhang, X. W. Lv and J. Li (2014). "Long noncoding RNAs: Novel insights into hepatocelluar carcinoma." <u>Cancer Lett</u> **344**(1): 20-27.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre,
H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." <u>Mol Cell</u> 38(4): 576-589.

Hillebrandt, S., H. E. Wasmuth, R. Weiskirchen, C. Hellerbrand, H. Keppeler, A. Werth, R. Schirin-Sokhan, G. Wilkens, A. Geier, J. Lorenzen, J. Kohl, A. M. Gressner, S. Matern and F. Lammert (2005). "Complement factor 5 is a quantitative trait gene that modifies liver fibrogenesis in mice and humans." Nat Genet **37**(8): 835-843.

Hines, R. N., S. B. Koukouritaki, M. T. Poch and M. C. Stephens (2008). "Regulatory polymorphisms and their contribution to interindividual differences in the expression of enzymes influencing drug and toxicant disposition." <u>Drug Metab Rev</u> **40**(2): 263-301.

Holloway, M. G., Y. Cui, E. V. Laz, A. Hosui, L. Hennighausen and D. J. Waxman (2007). "Loss of sexually dimorphic liver gene expression upon hepatocyte-specific deletion of Stat5a-Stat5b locus." <u>Endocrinology</u> **148**(5): 1977-1986.

Holloway, M. G., E. V. Laz and D. J. Waxman (2006). "Codependence of growth hormone-responsive, sexually dimorphic hepatic gene expression on signal transducer

and activator of transcription 5b and hepatic nuclear factor 4alpha." <u>Mol Endocrinol</u> **20**(3): 647-660.

Holloway, M. G., G. D. Miles, A. A. Dombkowski and D. J. Waxman (2008). "Liver-specific hepatocyte nuclear factor-4alpha deficiency: greater impact on gene expression in male than in female mouse liver." <u>Mol Endocrinol</u> **22**(5): 1274-1286.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." <u>Nucleic Acids Res</u> **37**(1): 1-13.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." <u>Nat Protoc</u> **4**(1): 44-57.

Huarte, M., M. Guttman, D. Feldser, M. Garber, M. J. Koziol, D. Kenzelmann-Broz, A. M. Khalil, O. Zuk, I. Amit, M. Rabani, L. D. Attardi, A. Regev, E. S. Lander, T. Jacks and J. L. Rinn (2010). "A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response." <u>Cell</u> **142**(3): 409-419.

Isensee, J., H. Witt, R. Pregla, R. Hetzer, V. Regitz-Zagrosek and P. R. Noppinger (2008). "Sexually dimorphic gene expression in the heart of mice and men." Journal of molecular medicine **86**(1): 61-74.

Jaffe, C. A., D. K. Turgeon, K. Lown, R. Demott-Friberg and P. B. Watkins (2002). "Growth hormone secretion pattern is an independent regulator of growth hormone actions in humans." <u>Am J Physiol Endocrinol Metab</u> **283**(5): E1008-1015.

Jakobsen, J. S., J. Waage, N. Rapin, H. C. Bisgaard, F. S. Larsen and B. T. Porse (2013). "Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries." <u>Genome Res</u> **23**(4): 592-603.

Jansson, J. O., S. Eden and O. Isaksson (1985). "Sexual dimorphism in the control of growth hormone secretion." Endocr Rev 6(2): 128-150.

Jansson, J. O., S. Eden and O. Isaksson (1985). "Sexual dimorphism in the control of growth hormone secretion." Endocrine reviews 6(2): 128-150.

Jansson, J. O., S. Ekberg, O. Isaksson, A. Mode and J. A. Gustafsson (1985). "Imprinting of growth hormone secretion, body growth, and hepatic steroid metabolism by neonatal testosterone." <u>Endocrinology</u> **117**(5): 1881-1889.

Kang, K., G. W. Robinson and L. Hennighausen (2013). "Comprehensive meta-analysis of Signal Transducers and Activators of Transcription (STAT) genomic binding patterns discerns cell-specific cis-regulatory modules." <u>BMC Genomics</u> 14: 4.

Karolchik, D., G. P. Barber, J. Casper, H. Clawson, M. S. Cline, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, A. S. Hinrichs, K. Learned, B. T. Lee, C. H. Li, B. J. Raney, B. Rhead, K. R. Rosenbloom, C. A. Sloan, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn and W. J. Kent (2014).
"The UCSC Genome Browser database: 2014 update." <u>Nucleic Acids Res</u> 42(Database issue): D764-770.

Kassam, I., L. Lloyd-Jones, A. Holloway, K. S. Small, B. Zeng, A. Bakshi, A. Metspalu, G. Gibson, T. D. Spector, T. Esko, G. W. Montgomery, J. E. Powell, J. Yang, P. M. Visscher and A. F. McRae (2016). "Autosomal genetic control of human gene expression does not differ across the sexes." <u>Genome Biol</u> **17**(1): 248.

Kaushik, K., V. E. Leonard, S. Kv, M. K. Lalwani, S. Jalali, A. Patowary, A. Joshi, V. Scaria and S. Sivasubbu (2013). "Dynamic expression of long non-coding RNAs (lncRNAs) in adult zebrafish." <u>PLoS One</u> **8**(12): e83616.

Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N. A. Furlotte, E. Eskin, C. Nellaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T. G. Belgard, P. L. Oliver, R. E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. van der Weyden, C. A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I. J. Jackson, A. Czechanski, J. A. Guerra-Assuncao, L. R. Donahue, L. G. Reinholdt, B. A. Payseur, C. P. Ponting, E. Birney, J. Flint and D. J. Adams (2011). "Mouse genomic variation and its effect on phenotypes and gene regulation." Nature 477(7364): 289-294.

Khalil, A. M., M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B. E. Bernstein, A. van Oudenaarden, A. Regev, E. S. Lander and J. L. Rinn (2009). "Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression." <u>Proc Natl Acad Sci U S A</u> **106**(28): 11667-11672.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley and S. L. Salzberg (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." <u>Genome Biol</u> **14**(4): R36.

Kim, T. K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman and M. E. Greenberg (2010). "Widespread transcription at neuronal activity-regulated enhancers." <u>Nature</u> **465**(7295): 182-187.

Kino, T., D. E. Hurt, T. Ichijo, N. Nader and G. P. Chrousos (2010). "Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor." <u>Sci Signal</u> **3**(107): ra8.

Klaassen, C. D. and L. M. Aleksunes (2010). "Xenobiotic, bile acid, and cholesterol transporters: function and regulation." <u>Pharmacological reviews</u> **62**(1): 1-96.

Ko, C. H. and J. S. Takahashi (2006). "Molecular components of the mammalian circadian clock." <u>Hum Mol Genet</u> **15 Spec No 2**: R271-277.

Kodama, S. and M. Negishi (2013). "PXR cross-talks with internal and external signals in physiological and pathophysiological responses." <u>Drug metabolism reviews</u> **45**(3): 300-310.

Krebs, C. J., S. Khan, J. W. MacDonald, M. Sorenson and D. M. Robins (2009). "Regulator of sex-limitation KRAB zinc finger proteins modulate sex-dependent and independent liver metabolism." <u>Physiol Genomics</u> **38**(1): 16-28.

Krebs, C. J., L. K. Larkins, S. M. Khan and D. M. Robins (2005). "Expansion and diversification of KRAB zinc-finger genes within a cluster including Regulator of sexlimitation 1 and 2." <u>Genomics</u> **85**(6): 752-761.

Krebs, C. J., L. K. Larkins, R. Price, K. M. Tullis, R. D. Miller and D. M. Robins (2003). "Regulator of sex-limitation (Rsl) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression." <u>Genes Dev</u> **17**(21): 2664-2674.

Krebs, C. J., D. C. Schultz and D. M. Robins (2012). "The KRAB zinc finger protein RSL1 regulates sex- and tissue-specific promoter methylation and dynamic hormone-responsive chromatin configuration." <u>Mol Cell Biol</u> **32**(18): 3732-3742.

Kundaje, A., W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P.
Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L.
D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfenning,
X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E.
Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J.
Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo,
M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A.
Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak,
N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M.
Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De
Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T.
McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A.
Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M.
Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang and M.

Kellis (2015). "Integrative analysis of 111 reference human epigenomes." <u>Nature</u> **518**(7539): 317-330.

Kung, J. T., D. Colognori and J. T. Lee (2013). "Long noncoding RNAs: past, present, and future." <u>Genetics</u> **193**(3): 651-669.

Kutter, C., S. Watt, K. Stefflova, M. D. Wilson, A. Goncalves, C. P. Ponting, D. T. Odom and A. C. Marques (2012). "Rapid turnover of long noncoding RNAs and the evolution of gene expression." <u>PLoS Genet</u> **8**(7): e1002841.

Langfelder, P. and S. Horvath (2008). "WGCNA: an R package for weighted correlation network analysis." <u>BMC Bioinformatics</u> **9**: 559.

Langfelder, P. and S. Horvath (2012). "Fast R Functions for Robust Correlations and Hierarchical Clustering." J Stat Softw **46**(11).

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." <u>Nat Methods</u> **9**(4): 357-359.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memoryefficient alignment of short DNA sequences to the human genome." <u>Genome Biol</u> **10**(3): R25.

Lau-Corona, D., A. Suvorov and D. J. Waxman (2017). "Feminization of male mouse liver by persistent growth hormone stimulation: Activation of sex-biased transcriptional networks and dynamic changes in chromatin states." <u>Mol Cell Biol</u>.

Laz, E. V., M. G. Holloway, C. S. Chen and D. J. Waxman (2007). "Characterization of three growth hormone-responsive transcription factors preferentially expressed in adult female liver." <u>Endocrinology</u> **148**(7): 3327-3337.

Leek, J. T., W. E. Johnson, H. S. Parker, A. E. Jaffe and J. D. Storey (2012). "The sva package for removing batch effects and other unwanted variation in high-throughput experiments." <u>Bioinformatics</u> **28**(6): 882-883.

Li, Z., M. Liu, L. Zhang, W. Zhang, G. Gao, Z. Zhu, L. Wei, Q. Fan and M. Long (2009). "Detection of intergenic non-coding RNAs expressed in the main developmental stages in Drosophila melanogaster." <u>Nucleic Acids Res</u> **37**(13): 4308-4314.

Li, Z., G. Tuteja, J. Schug and K. H. Kaestner (2012). "Foxa1 and Foxa2 are essential for sexual dimorphism in liver cancer." <u>Cell</u> **148**(1-2): 72-83.

Liao, Y., G. K. Smyth and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." <u>Bioinformatics</u> **30**(7): 923-930.

Lichanska, A. M. and M. J. Waters (2008). "How growth hormone controls growth, obesity and sexual dimorphism." <u>Trends Genet</u> **24**(1): 41-47.

Lichanska, A. M. and M. J. Waters (2008). "How growth hormone controls growth, obesity and sexual dimorphism." <u>Trends in genetics : TIG</u> **24**(1): 41-47.

Lin, M. F., I. Jungreis and M. Kellis (2011). "PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions." <u>Bioinformatics</u> **27**(13): i275-282.

Lindblad-Toh, K., M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P.
Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway,
M. Clamp, S. Gnerre, J. Alfoldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S.
Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D.
Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B. J. Raney, M. D.
Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody, J. Baldwin,
T. Bloom, C. W. Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K. C.
Worley, C. L. Kovar, D. M. Muzny, R. A. Gibbs, A. Cree, H. H. Dihn, G. Fowler, S.
Jhangiani, V. Joshi, S. Lee, L. R. Lewis, L. V. Nazareth, G. Okwuonu, J. Santibanez, W.
C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson, K. Delehaunty, D. Dooling, C.
Fronik, L. Fulton, B. Fulton, T. Graves, P. Minx, E. Sodergren, E. Birney, E. H.
Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J.
S. Pedersen, E. S. Lander and M. Kellis (2011). "A high-resolution map of human evolutionary constraint using 29 mammals." Nature **478**(7370): 476-482.

Ling, G., A. Sugathan, T. Mazor, E. Fraenkel and D. J. Waxman (2010). "Unbiased, genome-wide in vivo mapping of transcriptional regulatory elements reveals sex differences in chromatin structure associated with sex-specific liver gene expression." <u>Mol Cell Biol</u> **30**(23): 5531-5544.

Liu, W., K. Kulkarni and M. Hu (2013). "Gender-dependent differences in uridine 5'diphospho-glucuronosyltransferase have implications in metabolism and clearance of xenobiotics." <u>Expert Opin Drug Metab Toxicol</u> **9**(12): 1555-1569.

Lodato, N. J., T. Melia, A. Rampersaud and D. J. Waxman (2017). "Sex-Differential Responses of Tumor Promotion-Associated Genes and Dysregulation of Novel Long Noncoding RNAs in Constitutive Androstane Receptor-Activated Mouse Liver." <u>Toxicol Sci</u> **159**(1): 25-41.

Lowe, R., C. Gemma, V. K. Rakyan and M. L. Holland (2015). "Sexually dimorphic gene expression emerges with embryonic genome activation and is dynamic throughout development." <u>BMC Genomics</u> **16**: 295.

Lu, H., L. Zeng, B. Liang, X. Shu and D. Xie (2009). "High prevalence of coronary heart disease in type 2 diabetic patients with non-alcoholic fatty liver disease." <u>Arch Med Res</u> **40**(7): 571-575.

Luo, H., S. Sun, P. Li, D. Bu, H. Cao and Y. Zhao (2013). "Comprehensive characterization of 10,571 mouse large intergenic noncoding RNAs from whole transcriptome sequencing." <u>PLoS One</u> **8**(8): e70835.

MacLeod, J. N., N. A. Pampori and B. H. Shapiro (1991). "Sex differences in the ultradian pattern of plasma growth hormone concentrations in mice." <u>J Endocrinol</u> **131**(3): 395-399.

Mahmud, S. A., L. S. Manlove and M. A. Farrar (2013). "Interleukin-2 and STAT5 in regulatory T cell development and function." <u>JAK-STAT</u> **2**(1): e23154.

Matouk, I. J., N. DeGroot, S. Mezan, S. Ayesh, R. Abu-lail, A. Hochberg and E. Galun (2007). "The H19 non-coding RNA is essential for human tumor growth." <u>PLoS One</u> **2**(9): e845.

Matthews, B. and D. J. Waxman (2017). "Insulation of gene expression by subTAD loop structures." <u>Unpublished manuscript</u>.

Melia, T., P. Hao, F. Yilmaz and D. J. Waxman (2016). "Hepatic Long Intergenic Noncoding RNAs: High Promoter Conservation and Dynamic, Sex-Dependent Transcriptional Regulation by Growth Hormone." <u>Mol Cell Biol</u> **36**(1): 50-69.

Menet, J. S., J. Rodriguez, K. C. Abruzzi and M. Rosbash (2012). "Nascent-Seq reveals novel features of mouse circadian transcriptional regulation." <u>Elife</u> 1: e00011.

Meyer, R. D., E. V. Laz, T. Su and D. J. Waxman (2009). "Male-specific hepatic Bcl6: growth hormone-induced block of transcription elongation in females and binding to target genes inversely coordinated with STAT5." <u>Mol Endocrinol</u> **23**(11): 1914-1926.

Monti, S., P. Tamayo, J. Mesirov and T. Golub (2003). "Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data." <u>Machine Learning</u> **52**(1-2): 91-118.

Morgan, A. P., C. P. Fu, C. Y. Kao, C. E. Welsh, J. P. Didion, L. Yadgary, L. Hyacinth, M. T. Ferris, T. A. Bell, D. R. Miller, P. Giusti-Rodriguez, R. J. Nonneman, K. D. Cook, J. K. Whitmire, L. E. Gralinski, M. Keller, A. D. Attie, G. A. Churchill, P. Petkov, P. F. Sullivan, J. R. Brennan, L. McMillan and F. Pardo-Manuel de Villena (2015). "The Mouse Universal Genotyping Array: From Substrains to Subspecies." <u>G3 (Bethesda)</u> **6**(2): 263-279.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." <u>Nat Methods</u> 5(7): 621-628.

Mudunuri, U., A. Che, M. Yi and R. M. Stephens (2009). "bioDBnet: the biological database network." <u>Bioinformatics</u> **25**(4): 555-556.

Mugford, C. A. and G. L. Kedderis (1998). "Sex-dependent metabolism of xenobiotics." <u>Drug Metab Rev</u> **30**(3): 441-498.

Munger, S. C., N. Raghupathy, K. Choi, A. K. Simons, D. M. Gatti, D. A. Hinerfeld, K. L. Svenson, M. P. Keller, A. D. Attie, M. A. Hibbs, J. H. Graber, E. J. Chesler and G. A. Churchill (2014). "RNA-Seq alignment to individualized genomes improves transcript abundance estimates in multiparent populations." <u>Genetics</u> **198**(1): 59-73.

O'Leary, N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy and K. D. Pruitt (2016). "Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation." <u>Nucleic Acids Res</u> 44(D1): D733-745.

Odet, F., W. Pan, T. A. Bell, S. G. Goodson, A. M. Stevans, Z. Yun, D. L. Aylor, C. Y. Kao, L. McMillan, F. P. de Villena and D. A. O'Brien (2015). "The Founder Strains of the Collaborative Cross Express a Complex Combination of Advantageous and Deleterious Traits for Male Reproduction." <u>G3 (Bethesda)</u> **5**(12): 2671-2683.

Odom, D. T., R. D. Dowell, E. S. Jacobsen, W. Gordon, T. W. Danford, K. D. MacIsaac, P. A. Rolfe, C. M. Conboy, D. K. Gifford and E. Fraenkel (2007). "Tissue-specific transcriptional regulation has diverged significantly between human and mouse." <u>Nat Genet</u> **39**(6): 730-732.

Ohhata, T., M. Matsumoto, M. Leeb, S. Shibata, S. Sakai, K. Kitagawa, H. Niida, M. Kitagawa and A. Wutz (2015). "Histone H3 Lysine 36 Trimethylation Is Established over the Xist Promoter by Antisense Tsix Transcription and Contributes to Repressing Xist Expression." <u>Mol Cell Biol</u> **35**(22): 3909-3920.

Orom, U. A., T. Derrien, M. Beringer, K. Gumireddy, A. Gardini, G. Bussotti, F. Lai, M. Zytnicki, C. Notredame, Q. Huang, R. Guigo and R. Shiekhattar (2010). "Long noncoding RNAs with enhancer-like function in human cells." <u>Cell</u> **143**(1): 46-58.

Panzitt, K., M. M. Tschernatsch, C. Guelly, T. Moustafa, M. Stradner, H. M. Strohmaier, C. R. Buck, H. Denk, R. Schroeder, M. Trauner and K. Zatloukal (2007). "Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA." Gastroenterology **132**(1): 330-342.

Papp, A. C., J. K. Pinsonneault, D. Wang, L. C. Newman, Y. Gong, J. A. Johnson, C. J. Pepine, M. Kumari, A. D. Hingorani, P. J. Talmud, S. Shah, S. E. Humphries and W.

Sadee (2012). "Cholesteryl Ester Transfer Protein (CETP) polymorphisms affect mRNA splicing, HDL levels, and sex-dependent cardiovascular risk." PLoS One 7(3): e31930.

Peng, L., A. Paulson, H. Li, S. Piekos, X. He, L. Li and X. B. Zhong (2014). "Developmental programming of long non-coding RNAs during postnatal liver maturation in mice." <u>PLoS One</u> **9**(12): e114917.

Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R.
Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C.
Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein (2000). "Molecular portraits of human breast tumours." <u>Nature</u> 406(6797): 747-752.

Pincus, S. M., E. F. Gevers, I. C. Robinson, G. van den Berg, F. Roelfsema, M. L. Hartman and J. D. Veldhuis (1996). "Females secrete growth hormone with more process irregularity than males in both humans and rats." <u>Am J Physiol</u> **270**(1 Pt 1): E107-115.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn and J. Wysocka (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." <u>Nature</u> **470**(7333): 279-283.

Reich, M., T. Liefeld, J. Gould, J. Lerner, P. Tamayo and J. P. Mesirov (2006). "GenePattern 2.0." <u>Nat Genet</u> **38**(5): 500-501.

Reizel, Y., A. Spiro, O. Sabag, Y. Skversky, M. Hecht, I. Keshet, B. P. Berman and H. Cedar (2015). "Gender-specific postnatal demethylation and establishment of epigenetic memory." <u>Genes Dev</u> **29**(9): 923-933.

Renaud, H. J., J. Y. Cui, M. Khan and C. D. Klaassen (2011). "Tissue distribution and gender-divergent expression of 78 cytochrome P450 mRNAs in mice." <u>Toxicol Sci</u> **124**(2): 261-277.

Renaud, H. J., Y. J. Cui, H. Lu, X. B. Zhong and C. D. Klaassen (2014). "Ontogeny of hepatic energy metabolism genes in mice as revealed by RNA-sequencing." <u>PLoS One</u> **9**(8): e104560.

Rinn, J. L. and H. Y. Chang (2012). "Genome regulation by long noncoding RNAs." <u>Annu Rev Biochem</u> **81**: 145-166.

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal and H. Y. Chang (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." <u>Cell</u> **129**(7): 1311-1323.

Rinn, J. L., J. S. Rozowsky, I. J. Laurenzi, P. H. Petersen, K. Zou, W. Zhong, M. Gerstein and M. Snyder (2004). "Major molecular differences between mammalian sexes are involved in drug metabolism and renal function." <u>Dev Cell</u> **6**(6): 791-800.

Rinn, J. L. and M. Snyder (2005). "Sexual dimorphism in mammalian gene expression." <u>Trends Genet</u> **21**(5): 298-305.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." <u>Bioinformatics</u> **26**(1): 139-140.

Rogers, A. B., E. J. Theve, Y. Feng, R. C. Fry, K. Taghizadeh, K. M. Clapp, C. Boussahmain, K. S. Cormier and J. G. Fox (2007). "Hepatocellular carcinoma associated with liver-gender disruption in male mice." <u>Cancer Res</u> **67**(24): 11536-11546.

Rosenbloom, K. R., J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn and W. J. Kent (2015). "The UCSC Genome Browser database: 2015 update." <u>Nucleic Acids Res</u> **43**(Database issue): D670-681.

Ruggieri, A., C. Barbati and W. Malorni (2010). "Cellular and molecular mechanisms involved in hepatocellular carcinoma gender disparity." Int J Cancer **127**(3): 499-504.

Ruiz, M. L., A. D. Mottino, V. A. Catania and M. Vore (2013). "Hormonal regulation of hepatic drug biotransformation and transport systems." <u>Comprehensive Physiology</u> **3**(4): 1721-1740.

Sakakibara, M., Y. Uenoyama, S. Minabe, Y. Watanabe, C. Deura, S. Nakamura, G. Suzuki, K. Maeda and H. Tsukamura (2013). "Microarray analysis of perinatal-estrogeninduced changes in gene expression related to brain sexual differentiation in mice." <u>PLoS</u> <u>One</u> **8**(11): e79437.

Sauvageau, M., L. A. Goff, S. Lodato, B. Bonev, A. F. Groff, C. Gerhardinger, D. B. Sanchez-Gomez, E. Hacisuleyman, E. Li, M. Spence, S. C. Liapis, W. Mallard, M. Morse, M. R. Swerdel, M. F. D'Ecclessis, J. C. Moore, V. Lai, G. Gong, G. D. Yancopoulos, D. Frendewey, M. Kellis, R. P. Hart, D. M. Valenzuela, P. Arlotta and J. L. Rinn (2013). "Multiple knockout mouse models reveal lincRNAs are required for life and brain development." <u>Elife</u> **2**: e01749.

Schein, A., S. Zucchelli, S. Kauppinen, S. Gustincich and P. Carninci (2016). "Identification of antisense long noncoding RNAs that function as SINEUPs in human cells." <u>Sci Rep</u> **6**: 33605. Schmidt, C. W. (2015). "Diversity outbred: a new generation of mouse model." <u>Environ</u> <u>Health Perspect</u> **123**(3): A64-67.

Schmidt, D., M. D. Wilson, B. Ballester, P. C. Schwalie, G. D. Brown, A. Marshall, C. Kutter, S. Watt, C. P. Martinez-Jimenez, S. Mackay, I. Talianidis, P. Flicek and D. T. Odom (2010). "Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding." <u>Science</u> **328**(5981): 1036-1040.

Schmitt, B. M., K. L. Rudolph, P. Karagianni, N. A. Fonseca, R. J. White, I. Talianidis, D. T. Odom, J. C. Marioni and C. Kutter (2014). "High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface." <u>Genome Res</u> **24**(11): 1797-1807.

Sedek, M., L. M. van der Velden and G. J. Strous (2014). "Multimeric growth hormone receptor complexes serve as signaling platforms." J Biol Chem **289**(1): 65-73.

Shapiro, B. H., A. K. Agrawal and N. A. Pampori (1995). "Gender differences in drug metabolism regulated by growth hormone." <u>Int J Biochem Cell Biol</u> **27**(1): 9-20.

Shen, L., N. Y. Shao, X. Liu, I. Maze, J. Feng and E. J. Nestler (2013). "diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates." <u>PLoS One</u> **8**(6): e65598.

Shimizu, I., N. Kohno, K. Tamaki, M. Shono, H. W. Huang, J. H. He and D. F. Yao (2007). "Female hepatology: favorable role of estrogen in chronic liver disease with hepatitis B virus infection." <u>World J Gastroenterol</u> **13**(32): 4295-4305.

Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller and D. Haussler (2005). "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes." <u>Genome Res</u> **15**(8): 1034-1050.

Slater, G. S. and E. Birney (2005). "Automated generation of heuristics for biological sequence comparison." <u>BMC Bioinformatics</u> **6**: 31.

Smoot, M. E., K. Ono, J. Ruscheinski, P. L. Wang and T. Ideker (2011). "Cytoscape 2.8: new features for data integration and network visualization." <u>Bioinformatics</u> 27(3): 431-432.

Smyth, G. K. (2004). "Linear models and empirical bayes methods for assessing differential expression in microarray experiments." <u>Stat Appl Genet Mol Biol</u> **3**: Article3.

Soldaini, E., S. John, S. Moro, J. Bollenbacher, U. Schindler and W. J. Leonard (2000). "DNA binding site selection of dimeric and tetrameric Stat5 proteins reveals a large repertoire of divergent tetrameric Stat5a binding sites." <u>Mol Cell Biol</u> **20**(1): 389-401. Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning and A. L. Borresen-Dale (2001). "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." <u>Proc Natl</u> Acad Sci U S A **98**(19): 10869-10874.

Stamatoyannopoulos, J. A., M. Snyder, R. Hardison, B. Ren, T. Gingeras, D. M. Gilbert, M. Groudine, M. Bender, R. Kaul, T. Canfield, E. Giste, A. Johnson, M. Zhang, G. Balasundaram, R. Byron, V. Roach, P. J. Sabo, R. Sandstrom, A. S. Stehling, R. E. Thurman, S. M. Weissman, P. Cayting, M. Hariharan, J. Lian, Y. Cheng, S. G. Landt, Z. Ma, B. J. Wold, J. Dekker, G. E. Crawford, C. A. Keller, W. Wu, C. Morrissey, S. A. Kumar, T. Mishra, D. Jain, M. Byrska-Bishop, D. Blankenberg, B. R. Lajoie, G. Jain, A. Sanyal, K. B. Chen, O. Denas, J. Taylor, G. A. Blobel, M. J. Weiss, M. Pimkin, W. Deng, G. K. Marinov, B. A. Williams, K. I. Fisher-Aylor, G. Desalvo, A. Kiralusha, D. Trout, H. Amrhein, A. Mortazavi, L. Edsall, D. McCleary, S. Kuan, Y. Shen, F. Yue, Z. Ye, C. A. Davis, C. Zaleski, S. Jha, C. Xue, A. Dobin, W. Lin, M. Fastuca, H. Wang, R. Guigo, S. Djebali, J. Lagarde, T. Ryba, T. Sasaki, V. S. Malladi, M. S. Cline, V. M. Kirkup, K. Learned, K. R. Rosenbloom, W. J. Kent, E. A. Feingold, P. J. Good, M. Pazin, R. F. Lowdon and L. B. Adams (2012). "An encyclopedia of mouse DNA elements (Mouse ENCODE)." <u>Genome Biol</u> 13(8): 418.

Stubbington, M. J., B. Mahata, V. Svensson, A. Deonarine, J. K. Nissen, A. G. Betz and S. A. Teichmann (2015). "An atlas of mouse CD4(+) T cell transcriptomes." <u>Biol Direct</u> **10**: 14.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **102**(43): 15545-15550.

Sueyoshi, T., R. Kobayashi, K. Nishio, K. Aida, R. Moore, T. Wada, H. Handa and M. Negishi (1995). "A nuclear factor (NF2d9) that binds to the male-specific P450 (Cyp 2d-9) gene in mouse liver." <u>Mol Cell Biol</u> **15**(8): 4158-4166.

Sugathan, A. and D. J. Waxman (2013). "Genome-wide analysis of chromatin states reveals distinct mechanisms of sex-dependent gene regulation in male and female mouse liver." <u>Mol Cell Biol</u> **33**(18): 3594-3610.

Sun, H., J. Wu, P. Wickramasinghe, S. Pal, R. Gupta, A. Bhattacharyya, F. J. Agosto-Perez, L. C. Showe, T. H. Huang and R. V. Davuluri (2011). "Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq." <u>Nucleic Acids Res</u> **39**(1): 190-201. Svenson, K. L., D. M. Gatti, W. Valdar, C. E. Welsh, R. Cheng, E. J. Chesler, A. A. Palmer, L. McMillan and G. A. Churchill (2012). "High-resolution genetic mapping using the Mouse Diversity outbred population." <u>Genetics</u> **190**(2): 437-447.

Svenson, K. L., R. Von Smith, P. A. Magnani, H. R. Suetin, B. Paigen, J. K. Naggert, R. Li, G. A. Churchill and L. L. Peters (2007). "Multiple trait measurements in 43 inbred mouse strains capture the phenotypic diversity characteristic of human populations." J <u>Appl Physiol (1985)</u> **102**(6): 2369-2378.

Takahashi, K., I. Yan, H. Haga and T. Patel (2014). "Long noncoding RNA in liver diseases." <u>Hepatology</u> **60**(2): 744-753.

Tanaka, E. (1999). "Gender-related differences in pharmacokinetics and their clinical significance." J Clin Pharm Ther **24**(5): 339-346.

Tannenbaum, G. S. and J. B. Martin (1976). "Evidence for an endogenous ultradian rhythm governing growth hormone secretion in the rat." <u>Endocrinology</u> **98**(3): 562-570.

Tessari, P., A. Coracina, A. Cosma and A. Tiengo (2009). "Hepatic lipid metabolism and non-alcoholic fatty liver disease." <u>Nutrition, metabolism, and cardiovascular diseases :</u> <u>NMCD</u> **19**(4): 291-302.

Threadgill, D. W., D. R. Miller, G. A. Churchill and F. P. de Villena (2011). "The collaborative cross: a recombinant inbred mouse population for the systems genetic era." Ilar j **52**(1): 24-31.

Thurmann, P. A. and B. C. Hompesch (1998). "Influence of gender on the pharmacokinetics and pharmacodynamics of drugs." <u>Int J Clin Pharmacol Ther</u> **36**(11): 586-590.

Tojima, H., S. Kakizaki, Y. Yamazaki, D. Takizawa, N. Horiguchi, K. Sato and M. Mori (2012). "Ligand dependent hepatic gene expression profiles of nuclear receptors CAR and PXR." <u>Toxicol Lett</u> **212**(3): 288-297.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold and L. Pachter (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotechnol **28**(5): 511-515.

Tripathi, V., J. D. Ellis, Z. Shen, D. Y. Song, Q. Pan, A. T. Watt, S. M. Freier, C. F. Bennett, A. Sharma, P. A. Bubulya, B. J. Blencowe, S. G. Prasanth and K. V. Prasanth (2010). "The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation." <u>Mol Cell</u> **39**(6): 925-938.

Tryndyak, V., A. de Conti, T. Kobets, K. Kutanzi, I. Koturbash, T. Han, J. C. Fuscoe, J. R. Latendresse, S. Melnyk, S. Shymonyak, L. Collins, S. A. Ross, I. Rusyn, F. A. Beland and I. P. Pogribny (2012). "Interstrain differences in the severity of liver injury induced by a choline- and folate-deficient diet in mice are associated with dysregulation of genes involved in lipid metabolism." <u>FASEB J 26(11)</u>: 4592-4602.

Tsai, M. C., O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal and H. Y. Chang (2010). "Long noncoding RNA as modular scaffold of histone modification complexes." <u>Science</u> **329**(5992): 689-693.

Tyler, A. L., B. Ji, D. M. Gatti, S. C. Munger, G. A. Churchill, K. L. Svenson and G. W. Carter (2017). "Epistatic Networks Jointly Influence Phenotypes Related to Metabolic Disease and Gene Expression in Diversity Outbred Mice." <u>Genetics</u> **206**(2): 621-639.

Udy, G. B., R. P. Towers, R. G. Snell, R. J. Wilkins, S. H. Park, P. A. Ram, D. J. Waxman and H. W. Davey (1997). "Requirement of STAT5b for sexual dimorphism of body growth rates and liver gene expression." <u>Proc Natl Acad Sci U S A</u> **94**(14): 7239-7244.

Ulitsky, I., A. Shkumatava, C. H. Jan, H. Sive and D. P. Bartel (2011). "Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution." <u>Cell</u> **147**(7): 1537-1550.

van Nas, A., D. Guhathakurta, S. S. Wang, N. Yehya, S. Horvath, B. Zhang, L. Ingram-Drake, G. Chaudhuri, E. E. Schadt, T. A. Drake, A. P. Arnold and A. J. Lusis (2009). "Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks." <u>Endocrinology</u> **150**(3): 1235-1249.

Veldhuis, J. D., S. M. Anderson, N. Shah, M. Bray, T. Vick, A. Gentili, T. Mulligan, M. L. Johnson, A. Weltman, W. S. Evans and A. Iranmanesh (2001). "Neurophysiological regulation and target-tissue impact of the pulsatile mode of growth hormone secretion in the human." <u>Growth hormone & IGF research : official journal of the Growth Hormone</u> Research Society and the International IGF Research Society **11 Suppl A**: S25-37.

Veldhuis, J. D., S. M. Anderson, N. Shah, M. Bray, T. Vick, A. Gentili, T. Mulligan, M. L. Johnson, A. Weltman, W. S. Evans and A. Iranmanesh (2001). "Neurophysiological regulation and target-tissue impact of the pulsatile mode of growth hormone secretion in the human." <u>Growth Horm IGF Res</u> **11 Suppl A**: S25-37.

Veldhuis, J. D. and C. Y. Bowers (2003). "Human GH pulsatility: an ensemble property regulated by age and gender." <u>J Endocrinol Invest</u> **26**(9): 799-813.

Wang, J., X. Liu, H. Wu, P. Ni, Z. Gu, Y. Qiao, N. Chen, F. Sun and Q. Fan (2010). "CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer." <u>Nucleic Acids Res</u> **38**(16): 5366-5383. Wang, K. C., Y. W. Yang, B. Liu, A. Sanyal, R. Corces-Zimmerman, Y. Chen, B. R. Lajoie, A. Protacio, R. A. Flynn, R. A. Gupta, J. Wysocka, M. Lei, J. Dekker, J. A. Helms and H. Y. Chang (2011). "A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression." <u>Nature</u> **472**(7341): 120-124.

Wang, X., F. Magkos and B. Mittendorfer (2011). "Sex differences in lipid and lipoprotein metabolism: it's not just about sex hormones." <u>J Clin Endocrinol Metab</u> **96**(4): 885-893.

Wauthier, V., A. Sugathan, R. D. Meyer, A. A. Dombkowski and D. J. Waxman (2010). "Intrinsic sex differences in the early growth hormone responsiveness of sex-specific genes in mouse liver." <u>Mol Endocrinol</u> **24**(3): 667-678.

Wauthier, V. and D. J. Waxman (2008). "Sex-specific early growth hormone response genes in rat liver." <u>Mol Endocrinol</u> **22**(8): 1962-1974.

Waxman, D. J., G. A. Dannan and F. P. Guengerich (1985). "Regulation of rat hepatic cytochrome P-450: age-dependent expression, hormonal imprinting, and xenobiotic inducibility of sex-specific isoenzymes." <u>Biochemistry</u> **24**(16): 4409-4417.

Waxman, D. J. and M. G. Holloway (2009). "Sex differences in the expression of hepatic drug metabolizing enzymes." <u>Mol Pharmacol</u> **76**(2): 215-228.

Waxman, D. J. and C. O'Connor (2006). "Growth hormone regulation of sex-dependent liver gene expression." <u>Mol Endocrinol</u> **20**(11): 2613-2629.

Waxman, D. J., N. A. Pampori, P. A. Ram, A. K. Agrawal and B. H. Shapiro (1991). "Interpulse interval in circulating growth hormone patterns regulates sexually dimorphic expression of hepatic cytochrome P450." <u>Proc Natl Acad Sci U S A</u> **88**(15): 6868-6872.

Waxman, D. J., P. A. Ram, S. H. Park and H. K. Choi (1995). "Intermittent plasma growth hormone triggers tyrosine phosphorylation and nuclear translocation of a liver-expressed, Stat 5-related DNA binding protein. Proposed role as an intracellular regulator of male-specific liver gene transcription." <u>The Journal of biological chemistry</u> **270**(22): 13262-13270.

Waxman, D. J., P. A. Ram, S. H. Park and H. K. Choi (1995). "Intermittent plasma growth hormone triggers tyrosine phosphorylation and nuclear translocation of a liver-expressed, Stat 5-related DNA binding protein. Proposed role as an intracellular regulator of male-specific liver gene transcription." J Biol Chem **270**(22): 13262-13270.

Xie, C., J. Yuan, H. Li, M. Li, G. Zhao, D. Bu, W. Zhu, W. Wu, R. Chen and Y. Zhao (2014). "NONCODEv4: exploring the world of long non-coding RNA genes." <u>Nucleic Acids Res</u> **42**(Database issue): D98-103.
Yang, L., J. E. Froberg and J. T. Lee (2014). "Long noncoding RNAs: fresh perspectives into the RNA world." <u>Trends in biochemical sciences</u> **39**(1): 35-43.

Young, R. S., A. C. Marques, C. Tibbit, W. Haerty, A. R. Bassett, J. L. Liu and C. P. Ponting (2012). "Identification and properties of 1,119 candidate lincRNA loci in the Drosophila melanogaster genome." <u>Genome Biol Evol</u> **4**(4): 427-442.

Zanger, U. M. and M. Schwab (2013). "Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation." <u>Pharmacol Ther</u> **138**(1): 103-141.

Zanger, U. M. and M. Schwab (2013). "Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation." <u>Pharmacology & therapeutics</u> **138**(1): 103-141.

Zaret, K. S. and J. S. Carroll (2011). "Pioneer transcription factors: establishing competence for gene expression." <u>Genes Dev</u> **25**(21): 2227-2241.

Zhang, L., F. Yang, J. H. Yuan, S. X. Yuan, W. P. Zhou, X. S. Huo, D. Xu, H. S. Bi, F. Wang and S. H. Sun (2013). "Epigenetic activation of the MiR-200 family contributes to H19-mediated metastasis suppression in hepatocellular carcinoma." <u>Carcinogenesis</u> **34**(3): 577-586.

Zhang, X., H. Li, J. C. Burnett and J. J. Rossi (2014). "The role of antisense long noncoding RNA in small RNA-triggered gene activation." <u>Rna</u> **20**(12): 1916-1928.

Zhang, Y., K. Klein, A. Sugathan, N. Nassery, A. Dombkowski, U. M. Zanger and D. J. Waxman (2011). "Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease." <u>PLoS One</u> **6**(8): e23506.

Zhang, Y., E. V. Laz and D. J. Waxman (2012). "Dynamic, sex-differential STAT5 and BCL6 binding to sex-biased, growth hormone-regulated genes in adult mouse liver." <u>Mol</u> <u>Cell Biol</u> **32**(4): 880-896.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu (2008). "Model-based analysis of ChIP-Seq (MACS)." <u>Genome Biol</u> **9**(9): R137.

Zhao, J., T. K. Ohsumi, J. T. Kung, Y. Ogawa, D. J. Grau, K. Sarma, J. J. Song, R. E. Kingston, M. Borowsky and J. T. Lee (2010). "Genome-wide identification of polycomb-associated RNAs by RIP-seq." <u>Mol Cell</u> **40**(6): 939-953.

Zhao, Y., H. Li, S. Fang, Y. Kang, W. Wu, Y. Hao, Z. Li, D. Bu, N. Sun, M. Q. Zhang and R. Chen (2016). "NONCODE 2016: an informative and valuable data source of long non-coding RNAs." <u>Nucleic Acids Res</u> **44**(D1): D203-208.

## **CURRICULUM VITAE**

