University of Kentucky

UKnowledge

University of Kentucky Doctoral Dissertations | Graduate School

2011

ANALYSIS OF DIFFERENTIAL GENE EXPRESSION AND ALTERNATIVE SPLICING IN THE LIVER AND GASTROINTESTINAL TRACT IN THE LACTATING RAT

Antony Thomas Athippozhy
University of Kentucky, a.athippozhy@insightbb.com

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Abstract of Dissertation

Antony Thomas Athippozhy

The Graduate School

University of Kentucky

2011

ANALYSIS OF DIFFERENTIAL GENE EXPRESSION AND ALTERNATIVE SPLICING
IN THE LIVER AND GASTROINTESTINAL TRACT IN THE LACTATING RAT

_____

Abstract of Dissertation

_____

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
Graduate Center of Toxicology at the University of Kentucky

By

Antony Thomas Athippozhy

Lexington, Kentucky

Codirectors: Dr. Mary Vore, Professor of Toxicology

and Dr. Arnold Stromberg, Professor of Statistics

Lexington, Kentucky

Abstract

ANALYSIS OF DIFFERENTIAL GENE EXPRESSION AND ALTERNATIVE SPLICING IN THE LIVER AND GASTROINTESTINAL TRACT IN THE LACTATING RAT

Rat exon microarrays were utilized to detect changes in mRNA expression and alternative splicing in the liver, duodenum, jejunum, and ileum of the lactating rat when compared to age-matched virgin controls. Analysis of data at the level of gene expression revealed differential expression of genes involved in cholesterol biosynthesis in each tissue examined, suggesting increased Sterol Response Element Binding Protein activity. We also detected decreased mRNA from components of the T-cell signaling pathway in the jejunum and ileum. We characterized expression of solute carrier and adenosine triphosphate binding cassette proteins. In addition to characterizing genes by pathway, we have also grouped genes based on their pattern of expression to identify important genes. Amongst genes upregulated in all tissues was Slc39a4, which is a critical transporter in the absorption of zinc in enterocytes. Alternative splicing analysis detected a substantial amount of alternative splicing in the ileum compared to other tissues. In addition, in the liver Abcg8, a protein that functions as a heterodimer to export cholesterol in the bile, shows differential splicing in the liver, but not in other tissues.   We also detected differential expression of Ugt1a6 in the liver based on usage of an alternative first exon, which is consistent with altered protein levels observed previously. Differential splicing also appears to occur in Ace2 in the ileum, which could have consequences on the renin-angiotensin pathway.

Multimedia formats: .jpeg, pdf

KEYWORDS: Liver, small intestine, microarray, lactation, cholesterol

___Antony Athippozhy___

_____06/07/2011_____

ANALYSIS OF DIFFERENTIAL GENE EXPRESSION AND ALTERNATIVE SPLICING
IN THE LIVER AND GASTROINTESTINAL TRACT IN THE LACTATING RAT

By

Antony Thomas Athippozhy

_____Arnold J. Stromberg_____
(Co-director of dissertation)

_____Mary M. Vore_____
(Co-director of dissertation)

_____David K. Orren_____
(Director of Graduate Studies)

_____06/07/2011_____
(Date)

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name   Date

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Dissertation

Antony Thomas Athippozhy

The Graduate School

University of Kentucky

2011

ANALYSIS OF DIFFERENTIAL GENE EXPRESSION AND ALTERNATIVE SPLICING
IN THE LIVER AND GASTROINTESTINAL TRACT IN THE LACTATING RAT

_____
Dissertation
_____

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
Graduate Center for Toxicology
at the University of Kentucky

By

Antony Thomas Athippozhy

Lexington, Kentucky

Co-Directors:v Dr. Mary Vore , Professor of Toxicology

and Dr. Arnold Stromberg, Professor of Statistics

Lexington, Kentucky

ACKNOWLEDGEMENTS

Table of Contents

List of Tables

List of Figures

List of Files

Chapter 1

Physiology and Endocrinology of Lactating Animals

Introduction

Lactation is a time of increased energy demand, as lactating mothers must

provide nutrients for both themselves and their offspring. In rats, who have large

litters, energy demand is increased four to five fold, while food intake increases

two to three fold [1-3]. Several changes in hormone levels occur to stimulate the

physiologic changes that take place during lactation. Prolactin (Prl) is critical in

the development of the mammary gland and stimulates milk secretion [4-7], and

its levels are increased in lactation. Thyroid hormone, leptin, and insulin serum

levels are all decreased, possibly to maintain energy balance within the lactating

animal by decreasing its energy expenditure. Amongst changes seen in lactating

rats are increased liver size [8] , growth of the small intestine [1], increased

cholesterol synthesis [9], and increased hydrophobicity and size of the bile acid

pool [10]. Growth of visceral tissue and improved cholesterol synthesis are also

observed in bovines [11, 12].

In this study, we analyzed the liver, duodenum, jejunum, and ileum of lactating

dams and compared these tissues against the same tissues in virgin controls.

The liver is the primary site of drug metabolism, and is also responsible for the production of the various components of bile, which is important for lipid absorption and the absorption of lipid soluble vitamins. The small intestine is divided into three parts, the duodenum, jejunum, and ileum. The three parts have different physiologic properties, and therefore were analyzed as three different tissues. The duodenum, jejunum, and ileum as whole tissues also include many different cell types. The duodenum includes Brunner's glands, which are mucuous secretory cells, while the ileum contains lymph nodules known as Peyer's patches [13]. Gene expression of transporters differs between parts of the small intestine. For example, Abcc3 expression increases substantially in the ileum [14], while Abcc2 expression is high in the duodenum and is lower in the ileum [15-17]. A review of the expression of Solute carrier (Slc) and Adenosine triphosphate binding cassette (Abc) proteins in the parts of the small intestine can be found in Oostendorp et al [15]. These transporters can control oral availability of various drugs.

Thyroid Hormone in the Lactating Animal

Thyroid hormone comes in multiple forms, with 3,5,3′-L-triiodothyronine (T3) being the form that is considered biologically active [18]. Serum concentrations of thyroid hormone are decreased in lactating rats [19]. Thyroid hormone is required for milk secretion [7, 20, 21], and serum thyroid levels are below optimal for maximal secretion [7, 21]. 5' Monodeiodinase activity is increased in the mammary tissue, which allows for the mammary gland to locally produce active

thyroid hormone, permitting the mammary gland to meet energy demands, while conserving energy in other tissues [22]. The decrease in serum thyroid levels is likely to be in order to conserve energy, as thyroid hormone is strongly associated with energy metabolism. The actions of the thyroid hormone are divided into "genomic" effects and "non-genomic" effects. There are multiple forms of the thyroid receptor, which mediates the genomic effects of thyroid hormone, and are produced by two genes, *Thra* and *Thrb*. In the liver and small intestine, *Thrb* is the dominant form of the receptor [23]. Microarray technology has been used to characterize how thyroid hormone influences gene expression in cultured human fibroblasts [24], mouse brain [25], mouse liver [26-29], mouse osteoblasts [30], and rat liver [31]. Among genes regulated in human fibroblasts in response to T3 treatment is Kruppel Like Factor 9 (KLF9, also known as BTE binding protein or BTEB1). Knockout of Klf9 in mice results in shorter intestinal villi [32]. However, Klf9 is a transcriptional repressor and may also be involved in inhibiting growth [33]. In addition, thyroid hormone is associated with mitochondriogenesis [18]. Mitochondria are the major site of energy production within a cell, and increasing the number of mitochondria aids in a cell's ability to produce energy. Truncated variations of isoform one of *Thra* and isoform one of *Thrb* (TRα1 and TRβ1), have been shown to be specifically imported into mitochondria, suggesting a mechanism for thyroid hormone to control mitochondrial gene expression [34-36].

Nongenomic interactions of T3 are created through interactions with various proteins. One pathway that is influenced by a direct protein interaction is the PI3K/AKT pathway. T3 interacts directly with PI3K, which leads to phosphorylation of AKT [37, 38]. The PI3Ks are a group of lipid kinases that are subdivided into three classes based on sequence homology [39]. These enzymes catalyze the conversion of phosphatidyl inositol (4,5) bisphosphate (PIP2) into phosphatidyl inositol (3,4,5) triphosphate (PIP3) [40]. PIP3 can then be converted back into PIP2 by phosphorylates phosphatase and tensin homolog (PTEN), or PIP3 can phosphorylate the kinase 3'-Phosphoinositide-dependent protein kinase 1 (PDK1) [39, 40]. PDK1 in turn phosphorylates the protein product of v-akt murine thymoma viral oncogene homolog (AKT) [39, 41]. AKT in turn controls a large number of cellular responses, including phosphorylating the forkhead box transcription factors (FOXOs) and the mechanistic target of rapamycin (mTOR) when it is part of mTOR complex 1 (mTORC1) [41, 42]. Both of these processes control the ability of the cell to mobilize energy, the former by deactivating transcription factors, and the latter by inhibiting the ability of eukaryotic initiation factor 4E binding protein 1 (4EBP1) to bind to and inhibit eukaryotic initiation factor 4E. Decreased 4EBP1 binding allows for additional protein translation. The FOXO transcription factors control the expression of phophoenolpyruvate carboxykinase (PEPCK) and glucose-6-phosphatase (G6P), two of the key enzymes for gluconeogenesis [43, 44].

Leptin and insulin in the Lactating Animal

4

Leptin is a cytokine secreted by adipose tissue and aids in regulating dietary intake by acting on leptin receptors in the hypothalamus [45]. In general, high leptin levels inhibit dietary intake. During lactation, leptin levels are low, encouraging nutrient uptake through increased feeding [19]. Reducing leptin activity reduces Adenosine Monophosphate Activated Kinase (AMPK) activity, decreasing β-oxidation. AMPK is a major energy sensor for cells and plays a major role in control of several metabolic pathways, including downregulating gluconeogenesis in the liver, decreasing cholesterol synthesis by phosphorylating 3-Hydroxy-3-methylglutaryl Coenzyme A reductase (Hmgcr), controlling glucose uptake in muscle by upregulating Glut4, and improving fatty acid oxidation and decreasing fatty acid biosynthesis by inhibiting acetyl-CoA carboxylase (Acc), the rate limiting step of fatty acid synthesis [46, 47]. AMPK is an enzyme composed of a combination of three subunits [48, 49]. The completed enzyme functions as an energy sensor for the cell and is regulated by the AMP/ATP ratio in the cell as well as multiple hormone signals [47]. Reduced AMPK activity also encourages fatty acid synthesis, including stimulation of Acc, [46], through decreased phosphorylation [47]. AMPK downregulates gluconeogenesis by decreasing expression of PEPCK and G6P [47]. AMPK can also suppress Ppary coactivator 1α (Ppargc1α) by decreasing CRE-binding protein-regulated transcription coactivator 2 (CRTC2) activity [47, 50]. AMPK activity overlaps with several other signaling pathways, including those of insulin and adiponectin . Decreased AMPK activity also encourages cholesterol

5

synthesis through decreased phosphorylation of 3-hydroxy-3-methylglutaryl-CoA reductase, which is the rate limiting step of cholesterol synthesis [47]. AMPK decreases ACC transcription through decreasing the activities of the transcription factors SREBP-1c and carbohydrate responsive element binding protein (ChREBP) [47]. Insulin, which also shows decreased levels in lactation, decreases expression of the leptin receptor [51]. In cows, expression of the short and long forms of hepatic leptin receptor increases 40% during the transition from late pregnancy to early lactation [51]. Leptin mediates many of its downstream signaling through JAK2/STAT3 activation [52].

Prolactin in the Lactating Animal

Prolactin is a hormone secreted from lactotrophic cells in the pituitary gland [53, 54], and is also secreted from the mammary epithelium, placenta, brain, and immune system [53]. It is associated with mammary gland development and milk secretion. It is closely related to growth hormone and placental lactogen, which are believed to all have been derived from a common ancestral gene [54, 55]. Prl and placental lactogen bind to the prolactin receptor (Prlr), which has a long form, an intermediate form, and a short form in the rat [54, 56], while growth hormone has its own receptor [4]. Homodimerization of the Prlr in turn triggers a JAK2/STAT5 signaling cascade, which is considered the canonical mechanism for Prl signaling [4, 54]. Two JAK2 proteins phosphorylate each other and the activated JAK2 proteins in turn phosphorylate STAT5a. STAT5a dimerizes and

transports to the nucleus, where it functions as a transcription factor and regulates transcription

Hormones that have altered serum levels in lactation influence alternative splicing

The decrease in insulin and leptin levels and the increase in Prl signaling during lactation may lead to changes in pathways that ultimately control alternative splicing of genes. All three of these hormones lead to downstream PI3K signaling through different receptors [52, 57-61]. PI3k signaling in turn can lead to downstream phosphorylation of several serine-arginine (SR) proteins. SR proteins control splicing events through their interactions with RNA, allowing recruitment of the ribonuclear proteins that perform the splicing reaction [62-64]. SR proteins require phosphorylation to recognize the potential splice site, and dephosphorylation for the splicing reaction to be catalyzed [62, 65, 66]. Insulin signaling can trigger differential splicing of protein kinase C β II through the PI3K pathway and ultimately through differential phosphorylation of SRp40, a member of the SR family of proteins that are often involved in the regulation of alternative splicing [67]. Likewise, Prlr activation leads to differential splicing of neuronal nitric oxide synthase (nNOS) in the rat anterior pituitary cell line GH_3 [68]. Consequently, differential alternative splicing through differential phosphorylation of SR proteins downstream of PI3k signaling seems likely in lactating animals.

Energy Demands of Lactating animals

During lactation, many species are in negative energy balance [69]. Maintaining a negative energy balance can produce a state where the animal is conserving energy. Both thyroid and insulin, hormones that encourage energy expenditure, have decreased serum levels in lactation, which is consistent with an energy-poor state [3, 19, 70]. This negative energy balance has been proposed to be causative of the hypoleptinemia seen in lactation [69] and is proposed to be responsible for the increase in food intake during lactation [69]. However, this explanation does not fully account for changes in food intake or in leptin levels. In addition to energy demands, the suckling response has an influence on the diurnal regulation of serum leptin levels [69]. Also, although leptin levels may be reduced to compensate for negative energy balance, dietary intake does not increase sufficiently to offset negative energy balance, as treatment of lactating rats with exogenous leptin only decreases their dietary intake by 20% [3, 69]. Increased dietary consumption is associated with increased liver and intestinal growth during lactation. In ruminants, several studies indicate an increase in liver and intestinal growth corresponding to increased caloric uptake. Prl was initially believed to be responsible for gut hypertrophy [71], as treatment with bromocriptine, a drug that blocks Prl release, decreases small intestinal growth. However, a later study directly investigating Prl showed that increased prolactin levels are insufficient for intestinal hypertrophy [72].

Zinc absorption in lactating animals

Zinc (Zn^{+2}) absorption is improved during lactation in both women [73-75] and rats [76]. Duodenums from lactating rats display an increased uptake of ^{65}Zn. Duodenal zinc absorption increases throughout pregnancy and lactation and returns to normal levels after weaning [76]. Zinc is an essential micronutrient involved in bone formation, with calcium, phosphorus and magnesium being the others [75]. In humans, zinc is the only one of these nutrients to continue to have an increased absorption rate after birth [75], although evidence exists for Prl improving intestinal absorption of calcium in the duodenum [77]. In lactating rats, Prl increases the duodenum's ability to uptake calcium through increased expression of multiple transporters, including transient receptor potential vanilloid family members five and six (Trpv5 and Trpv6) [77].

The Slc39 proteins are metal transporters. Of these, Slc39a4 has been described as the primary transporter for intestinal uptake of zinc, and failure to produce a functional form of this protein in humans leads to a disease state known as acrodermatitis enteropathica [78]. There is an alternative mechanism for zinc uptake. Although this mechanism is not well understood, the primary method for treatment of acrodermatitis enteropathica is a high zinc diet. Because lactating animals have improved zinc absorption, one of these mechanisms may be improved in the lactating dam. Mechanisms controlling expression of Slc39a4 are poorly understood, but another zinc transporter in the same family, Slc39a1,

is upregulated in response to Prl treatment in prostrate-derived cancer cell lines [78, 79].

Changes in the bile acid pool in lactating dams

The hepatocyte produces bile acids, which function in the intestine to solubilize cholesterol and other lipids and facilitate their absorption. The bile acid pool size and its hydrophobicity are increased in lactation. During a time course study over days of lactation, a statistically significant increase in the bile acid pool size is first noticed on day 10 of lactation [10]. A shift in the diurnal rhythm regulating Cyp7a1, the rate limiting step of bile acid synthesis, and a significant increase in Cyp7a1 transcription occurs at 16 h when compared against age matched virgin controls. No such increase was detected in Cyp27a1, suggesting that the primary mechanism for increased bile acid pool size is an increase in Cyp7a1 activity. The increase in transcription occurs alongside increased recruitment of LXRα to the Cyp7a1 promoter, possibly through increased availability of LXR substrate in the form of cholesterol [80]. Decreased repression of Cyp7a1 transcription plays a role in Cyp7a1 upregulation during lactation. Expression of Fibroblast growth factor 15 (Fgf15) is downregulated in the ileum in lactation, leading to decreased extracellular signal related kinase 1 and 2 (Erk1/2) phosphorylation, as shown by the 88% decrease in phosphorylated Erk1/2 in the liver of lactating rats [80].Coupled with the knowledge that in primary human hepatocytes, FGF19, the homolog of Fgf15, inhibits CYP7a1 through the ERK1/2

pathway [81], these data provide a likely mechanism by which Cyp7a1 expression is increased in the liver of the lactating dam.

Changes in bile acid transport in lactating dams

In addition to the increased production of bile acids, expression of the bile acid transporters Abcb11(Bsep), Slc10a2 (Asbt), and Slc10a1 (Ntcp), is also upregulated. Bile acids are synthesized in the liver, exported into bile through Abcb11, taken into the enterocyte by Asbt, transported into the portal circulation by Ostα/β, and then taken back up in the hepatocyte by Ntcp. Abcb11and Ntcp show a significantly increased level of protein relative to virgin controls on day 2 post partum as a response to Prl, as shown by treating rats with Prl and measuring resulting Ntcp and Abcb11 levels [82]. The increase in Ntcp mRNA and protein can be directly linked to increased JAK/STAT5 signaling at a Stat5 response element at -1237 to -758 bps in the Ntcp promoter and is mediated by the long form of the Prlr [83]. Ntcp shows a two fold increase in protein and a 1.7 fold increase in Vmax for transport of taurocholate at day two post-partum, and the increase in expression and protein is maintained for two weeks [82]. Regulation of Asbt appears to be posttranscriptional, as levels of Asbt mRNA are not different between control and post-partum day 14-21 rats despite increased Asbt protein expression [84]. Bile flow, which is generated by secretion of bile salts, is essential for the biliary secretion of cholesterol and many xenobiotics, and is increased in lactating rats [8, 85]. Bile flow, bile acid secretory rate, and

hepatic clearance of taurocholate in response to taurocholate infusion was shown to be mediated by Prl [8].

Increased cholesterol synthesis in lactating animals

Cholesterol biosynthesis is increased in both the small intestine and the liver during lactation. This was initially shown by an increase in incorporation of tritiated water into newly synthesized cholesterol [9]. At day 14 postpartum, a significant increase in cholesterol synthesis was detected per total organ in both the small intestine and liver in lactating rats when compared against nonlacting postpartum rats. 3-Hydroxy-3-methylglutaryl Coenzyme A reductase (Hmgcr) shows increased activity in the liver of lactating rats sacrificed at 10:00 and 14:30 on day 10 post-partum compared to age-matched virgin controls [86]. Because cholesterol biosynthetic enzymes are all sensitive to SREBP-2 activity [87, 88], one possible mechanism is that SREBP-2 activity is increased during lactation in these tissues. A similar trend has also been observed in the mammary gland [89]. Increased amounts of cholesterol are essential in lactation, due in part because it is needed for incorporation into the milk for growth and neural development in the pups. However, cholesterol is also required for bile acid synthesis, which is also increased in lactation. As indicated above, bile acids are required for efficient intestinal absorption of cholesterol, lipid, and lipid-soluble vitamins [90].

12

Cholesterol synthesis and lipogenesis are controlled by three transcription factors, known as Srebps, which are made by two genes, the Srebfs. The *Srebf1* gene codes for the transcription factors Srebp-1a and Srebp-1c, while *Srebf2* codes for the protein Srebp-2. The three proteins differ in their ability to regulate the processes of lipid synthesis and cholesterol biosynthesis, but there is overlap between the target genes [87, 88]. Srebp-1a is more closely associated with fatty acid synthesis, while Srebp-2 is more closely associated with cholesterol biosynthesis, as shown in mice expressing dominant positive forms of the two proteins [88]. The rate limiting steps of the two processes, Acetyl-CoA carboxylase (Acaca or Acc) for fatty acid synthesis, and HMGCR for cholesterol synthesis, are shown to be upregulated in mice expressing the dominant positive form of Srebp-1a and Srebp-1c, as well as downregulated in Srebp chaperone (Scap) knockout mice [87].

The Srebps share a common mechanism for activation [91]. Under conditions of sufficient cholesterol concentrations, the protein product of insulin stimulated gene (Insig) is bound to cholesterol, and keeps an Insig/Scap/Srebp complex sequestered in the endoplasmic reticulum. When not bound to cholesterol, Insig is degraded, freeing Scap to escort the Srebp to the Golgi where it is cleaved and activated. The N-terminal of the Srebp protein can then travel to the nucleus, where it functions as a transcription factor and stimulates the production of target

13

genes, including genes involved in fatty acid synthesis and/or genes involved in cholesterol biosynthesis [91].

Microarray analyses of various models have identified several potential downstream targets of Prl [53]. Amongst these targets is *Srebf1*, which codes for the Srebp1a and Srebp1c proteins. The protein products of Srebf1 are strongly associated with controlling fatty acid and cholesterol biosynthesis [87]. Srebf1 is consistently downregulated in three different models of Prl deficiency, indicating that Prl can control transcription of Srebf1 [53, 92]. Due to the increased need of the lactating dam for both fatty acids and cholesterol, the activity of all three Srebp proteins may be improved.

Changes in cholesterol and phospholipid transport into bile in lactation

Components of bile include cholesterol and phospholipids, in addition to bile acids. While Abcb11 transports bile acids into bile, Abcg5 and Abcg8 function as a heterodimer to transport cholesterol, and Abcb4 (Mdr2) transports phospholipids into bile. mRNA expression of Abcg5 and Abcg8 in liver decreases over 90% throughout lactation [93], and despite this loss of Abcg5/Abcg8 in expression in liver there is no change in the cholesterol or phospholipid concentration in bile during lactation. However, following infusion of taurocholate, increased secretion of cholesterol was impaired relative to controls

14

indicating a decreased coupling of taurocholate and cholesterol excretion into bile [93]. Western analysis does not detect Abcg8 protein in the livers of lactating rats [93]. The decreased presence of function Abcg5/Abcg8 dimer uncouples taurocholate and cholesterol transport into bile and could provide a mechanism for retention of cholesterol while the lactating dam is in a physiologic state that requires more cholesterol [93]

The Ugt1a Locus in Lactation

The UDP glucuronosyltransferase (Ugt) superfamily of proteins is responsible for transferring glucuronic acid to their respective substrates in order to form water soluble molecules that are often biologically inactive and are subsequently excreted into urine or bile [93]. The *Ugt1* gene associated with these proteins generates several possible proteins based on usage of an alternative first exon and common exons 2-5 [93]. In transgenic mice expressing the human form of the UGT1A locus, UGT1A4 and UGT1A6 expression are both elevated during lactation [93]. In rats, protein and mRNA levels of Ugt1a6 are increased in the liver, and its expression is also increased by treatment of rats with Prl, suggesting that Ugt1a6 expression is sensitive to Prl levels [94]. These data provide the rationale for studying expression of the alternative exons in the rat Ugt1a locus. The Affymetrix Rat Exon 1.0 ST array (Affymetrix, Santa Clara, CA) provides probes that interrogate these exons, and consequently allow for detection of differential expression of the various mRNAs produced by the Ugt1a

locus. Substrates of Ugt1a6 include phenols such as acetaminophen, and consequently Ugt1a6 is important in drug metabolism in the lactating animal [95].

Rationale

In light of the large number of physiologic changes that occur in lactating animals relative to virgin animals of the same species, characterization of every transporter, enzyme, hormone, and transcription factor that has altered expression during lactation would be inefficient, time consuming, and expensive. Characterization of the transcriptome of the lactating animal allows a better understanding of the ability of lactating animals to absorb and metabolize nutrients and respond to changes in hormone levels. Understanding the transcriptome can in turn lead to hypotheses regarding mechanisms of how lactating animals absorb key nutrients, synthesize important metabolites, and respond to altered levels of hormones in serum. Here, we utilized the Affymetrix Rat Exon 1.0ST exon chip to characterize mRNA expression and differential alternative splicing in tissues from day 10, 16 h lactating rats compared to age-matched virgin controls. Day 10, 16h lactating rats were chosen because this is the first time point at which a statistically significant increase in expression and activity of Cyp7a1, the rate limiting step of bile acid biosynthesis, is observed [10]. The microarray analyses should support our previous results regarding expression of key enzymes, such as Cyp7a1, in bile acid and cholesterol synthesis as well as transporters important for their transport into bile [10, 93], in

16

addition to providing insight into their mechanisms of regulation. Finally, new

information regarding expression and regulation of additional pathways important

for changes of biological importance in the lactating dam should be discovered.

Chapter 2

Analysis of Microarray Data

Microarray Design

While technology now exists for high throughput analysis of large biological datasets, biological systems are extremely complicated and require advanced techniques to mine and interpret the data. For example, microarray technology allows for the measurement of most RNAs within a given sample. Bioinformatics is the field of study in which "information sciences" such as computer science and statistics are utilized to interpret biological data.

Microarray technology is based on nucleotide hybridization methods. Oligonucleotide probes are utilized to detect the presence of cDNA synthesized from RNA samples and tagged with a fluorescent dye. Original microarrays utilized probes designed for the 3' ends of genes and were utilized for the estimation of overall gene expression. Newer designs can contain probes designed to span an entire transcript. These arrays also allow the estimation of differential alternative splicing between different experimental conditions. Affymetrix is a supplier of chips utilizing oligonucleotide probes synthesized in situ on a glass plate. For simplicity's sake, data analysis will be discussed with respect to the Affymetrix platform, but many of the same concepts are applicable to other platforms.

Pooling

Microarray chips can be expensive when compared to the cost of individual biological samples. Consequently, one method to circumvent the cost of a microarray experiment is to combine mRNA/cDNA from several samples to reduce the variance between individual chips. As more samples are pooled, the mRNA/cDNA concentration for a given gene approaches the population average for the new pooled sample, reducing the effect an individual rat would have on the measurement of gene expression. Power calculations demonstrate that increasing the number of samples pooled and loaded onto a chip can give similar power to that obtained by running a larger number of arrays. In an experiment in which p samples are mixed, the number of arrays needed to obtain a similar power in the absence of pooling is approximately 1/p, assuming a large number of samples [96]. Therefore, an experiment utilizing six chips, each containing mRNA from five individual samples would be comparable in statistical power to an experiment containing thirty individual samples.

Preanalysis Data Treatment

When fluorescence intensities are initially measured from the array, a number of biases exist that are independent of the biological sources of variance [97]. For example, a difference in lighting on individual chips would lead to a chip effect that could influence interpretation of the results. Consequently, a number of pre-analysis steps are required to convert the fluorescence intensities into data that have biological relevance. These steps usually include a background correction

step, a normalization step, and a transformation step, as described below. One commonly used method for performing these tasks is the Robust Multichip Averaging (RMA) method, although other techniques exist.

The first step in adjusting the data is background correction. Affymetrix originally utilized intentionally mismatched probes at the thirteenth nucleotide in a 25mer oligonucleotide probe to determine amounts of nonspecific binding. However, this method was found to not be effective, as the measured intensity of the mismatch probe was found to increase as the intensity of the perfect match probe increased [98], suggesting that the mismatch probe had specific affinity for the target sequence. One method to circumvent the usage of mismatch probes is to assume that the background is based on random error, and calculate it from the standard deviation of perfect match signals, as is done in RMA. The RMA background correction assumes that the observed signal is produced as the sum of the actual signal from the bound target and random error. The bound target signal is assumed to be exponential and the error is assumed to be normally distributed. A description of the normalization procedure can be found at http://bmbolstad.com/misc/ComputeRMAFAQ/ComputeRMAFAQ.html.

Normalization is performed in order to make each chip comparable by removing nonbiological effects and allows for measurements on separate chips to be comparable to each other. The RMA algorithm utilizes quantile normalization [99]. In this method, probe intensities are ranked for each chip. Then for each rank (e.g., the highest, 50th highest, 100th highest, etc), the average across all

chips is taken, and the corresponding values are all set to the average. This forces the distribution of measurements on all chips to be equal [100].

The final steps toward getting the data into a workable format are 1) log2 transformation of the intensities allowing for the use of parametric statistical tests by adjusting the distribution of the measurements, and 2) summarizing the intensity values of individual probes into a single measurement for the corresponding exon or gene. In RMA, Tukey's median polish [99] is utilized to summarize the data, but other methods such as Tukey's biweight have also been utilized [99].

Detection of Differential Gene Expression

Once the data between each individual microarray chip is comparable to data from the other chips, a method of analysis needs to be chosen. Early microarray experiments utilized a fold-change approach. The problem with this approach is that it only compares the means of a given gene's expression between two groups, and not the variance within those measurements. Consequently, this method does not allow for the calculation of the likelihood of a false positive. Correcting for the likelihood of false positives within the experiment is discussed in the "Multiple Testing in Microarray Data" section below.

Either a parametric or nonparametric approach can be chosen to identify differentially expressed genes. The nonparametric approach makes no assumptions about the distribution of the data, but at the loss of statistical power.

21

The parametric approach requires that the data being tested be approximately normally distributed. ANOVA and Kruskal-Wallis tests have been frequently used in microarray datasets to detect differential gene expression.

We have studied a system in which lactating rats were compared against age-matched virgin controls. We analyzed four tissues from each rat and pooled RNA from the same tissue from four rats within the same physiologic state (control or lactating). The pools of four rats were consistent between tissues. In other words, if rats 1,2,3, and four composed pool one in the liver, the same four rats would compose pool one in the duodenum, jejunum and ileum. This led to a two-way, mixed models, repeated measures experimental design. Mixed models approaches can be utilized when multiple samples are taken from the same subject. In this case, the four tissues were all taken from the same sets of four rats. This led to a design in which we had two fixed effects (tissue and physiologic state) and one random effect (the subject variable, which in this case is a set of four rats.) A mixed model is comparable to the corresponding linear model, except that the model is now expanded so that each term contains a matrix that accounts for the individual samples. Accounting for the presence of the individual subjects also generates a second error term that accounts for differences between unique samples and is based on the covariance matrix. This means that a method for approximating the covariance matrix must be applied. One method, known as compound symmetry, was to assume that all covariances are equal.

Multiple Testing in Microarray Data

Due to the large size of a microarray data set, multiple tests have been performed in order to detect changes in gene expression. Two types of error could be made, rejecting the null hypothesis (measurements are made from the same distribution) when it should not have been rejected (Type I error, or a false positive) and failing to reject the null hypothesis when it should have been (Type II error, or a false negative). Although the probability of any given test producing a type I error (alpha) is small, the probability that a false positive exists within the entire data set, (the family-wise error rate) may still be very high. There are several ways to correct for this. One of the earliest methods to correct for multiple testing was the Bonferroni procedure [101]. However, this test is extremely conservative, and leads to a very small list of differentially expressed genes. A second approach is to correct for the family-wise error rate (FWER) [101]. This method attempts to control the probability that a false positive appears anywhere in the dataset. However, because of the exploratory nature of many microarray studies, this test is still overly conservative, as many experimenters would be willing to accept the existence of a few false positives within the dataset. The false discovery rate (FDR) method attempts to fix the ratio of false positives to total positives in a given dataset [102]. The Benjamini-Hochberg procedure for calculating false discovery rates lets $p_1, p_2, \ldots p_m$ be the rank ordered p-values, where m is the total number of tests. [All hypotheses where $p_i < i\alpha/m$ are rejected [102], where i is the i'th ranked p value and α is the

23

cutoff value for the FDR.]   A common cutoff for the FDR is 0.05, but even this may be too conservative for some studies. An alternative to fixing the FDR is to fix a set p-value cutoff, and calculate what the FDR or proportion of false positives (PFP) is at the given p-value. The challenge in choosing an appropriate cutoff is that as the probability of a type I error decreases, the probability of a type II error increases. Note that although the Benjamini-Hochberg [102] method was the first described to control the FDR, other methods exist [101]. Here we set the p value cutoff threshold and calculated the PFP.

Detection of Trends and Pathways in A List of Differentially Expressed Genes

Once a list of differentially expressed genes has been generated, the experimenter must determine which genes are of particular interest. Given a list of the differentially expressed genes and a database that organizes these genes into various categories (for example, organizing genes by pathway or function), a test can be performed to identify whether a given category is overrepresented in this list of genes. Several methods exist for detecting overrepresented groups of differentially-expressed genes [103]. One method is to utilize a right-tailed Fisher's Exact test to identify the probability that a number equal to or higher than the number of genes actually detected within a group would appear by chance [104]. This method is utilized by DAVID [104, 105] and Ingenuity Pathways Analysis (IPA) (Ingenuity Systems, www.ingenuity.com). Because a test is performed for each category of genes, multiple testing is an issue, and a multiple

24

testing procedure such as the FDR can be applied to correct for the possibility of a high number of type I errors [104].

Analysis of differential alternative splicing provides many additional challenges. On an Affymetrix exon array, the probes on the array are divided into probesets, which correspond to regions of a transcript. For the purposes of analysis, the probesets are each treated as independent exons, although large exons are often assigned multiple probesets. Based on the source of information Affymetrix utilized to design the probes, Affymetrix assigns each probeset to a "confidence level". "Core" probesets are regions that are drawn from a curated database such as Refseq. "Extended" probesets are derived from expressed sequence tag (EST) databases.   "Full" probesets are derived from software predictions. There are 92,354 probesets at the Core confidence level on the Affymetrix Rat Exon 1.0 ST array. Consequently, a large number of exons could test positive by chance, and multiple testing is problematic. The Benjamini-Hochberg approach has been applied to address this issue [106], and is readily available in software packages, but to our knowledge, the multiple testing issue has not been addressed specifically for exon arrays.

Detection of Differential Alternative Splicing

After performing a data treatment algorithm such as RMA, data can be summarized at both the probeset and transcript level. Summarizing data at the probeset level permits analysis at the exon level. However, a large difference between intensities of a given probeset is not sufficient for identifying

differentially alternatively spliced transcripts. Although a specific probeset may display differential expression, the transcript itself may be differentially expressed. Consequently, determination for whether or not a specific treatment triggers differences in alternative splicing must take into account differences in mRNA expression.

Affymetrix arrays lack junction-specific probes, and therefore methods designed for detecting alternative splicing using this information is invalid. Three of the methods that do not take into account this information are Analysis of Splicing Variance (ANOSVA) [107], Microarray Dectection of Alternative Splicing (MiDAS) [97], and Finding Isoforms from Robust Multichip Analysis (FIRMA) [108].

Affymetrix proposed the MiDAS method as a means of detecting differential alternative splicing between multiple groups. In this method, probeset intensities are divided by transcript level intensity to create a gene-normalized intensity. Then, a statistical test (an ANOVA, if done as originally proposed by Affymetrix) is utilized to determine differences in gene-normalized intensities between treatment groups. The first problem to address is that without adequate filtering, every exon in the dataset is tested at least once for alternative splicing. This leads to a very high number of statistical tests, and the problem of multiple testing becomes substantial, especially if the treatment triggers a relatively small number of changes in differential splicing when compared to the entire dataset. The second is that the calculated transcript level intensity needs to be accurate, because the algorithm is dependent on using the transcript level estimate to

26

normalize all probesets within that transcript. This can be particularly

problematic in a gene that has a large number of splicing events, since the

transcript level measurement is summarized from the probeset level

measurements (even if in principle, the individual probes are treated separately

for the purposes of summarization.) Reports exist stating that the distribution of

p-values resulting from MiDAS are not normally distributed in the non-significant

range [109].

The ANOSVA method utilizes a linear model to determine if differences in

alternative splicing occur between groups. The model follows the form of

$y_{ijkl}=\mu+\alpha_i+\beta_j+\gamma_{ij}+error$, where the error is normally distributed around mean zero

and random, μ is the baseline intensity, α_i is the "exon effect", β_j is a "treatment

effect", and γ_{ij} is the interaction term between α_i and β_j. The interaction term

(treatment*exon) is utilized as a means of determining if a transcript is

differentially alternatively spliced between treatment groups [107]. A significant

interaction term is determined to be a positive test. This conclusion operates

under the assumption that if a probeset's intensity is not the sum of random error,

the treatment effect, and the individual exon effect (which can also be considered

as a probe affinity term) then the difference must be caused by differential

alternative splicing. A significant interaction is graphically visible by plotting

intensities as a function of exon for all treatment groups. In the absence of an

interaction term, the intensities for each group will be parallel when plotted as a

function of the individual probesets (Figure 2.1). This can be seen in that the

27

intensities of both control and lactation are parallel for Abcg5 in the ileum, but in

Ace2 there is a point where the two lines were not parallel, indicating possible

alternative splicing at that point. However, there are a number of situations

where events other than differential alternative splicing cause a significant

interaction effect. This will lead to false positives, and because the underlying

assumptions regarding what is defined as alternative splicing are violated, even a

transcript with a low interaction p-value may be a false positive under certain

circumstances. These circumstances are discussed below.

FIRMA is a method based on utilizing the RMA algorithm. The model

$(PM_{ik}) = c_i + \rho_k + \mathcal{C}_{ik}$, is utilized to calculate residuals for each individual probe,

where PM_{ik} is the measured intensity from the probe, c_i is the chip effect

(expression level), p_k is the probe effect (or probe affinity), and $\mathcal{C}_{ik}$ is the random

error [108]. The median residual (the difference from the mean) divided by the

standard deviation is used as a score to identify exons that are differentially

spliced. Unfortunately, a good guideline for a significance cutoff based on

FIRMA scores has not been developed [108].

A number of events can occur that will lead to false positives. The first is the

absence of either a gene that is not expressed in a treatment group, or an exon

that is expressed in no treatment groups. Because the methods for detection are

based on the relative expression of a probeset relative to the gene, the absence

of a probeset in both groups or the absence of the gene in one group lead to

false assumptions in the algorithm. In order to filter out poorly expressed

28

probesets, Affymetrix has developed the Detection Above Background Algorithm (DABG) algorithm [110], as exon arrays do not have single nucleotide mismatch probes. In this method, background probes of random sequences of the 26 possible GC contents for a 25-mer oligonucleotide sequences are measured. These probes are used as a background for probes of the same GC content. Probesets with a DABG p-value of less than 0.05 can be considered present at above background levels, although making the threshold more stringent is possible. No method for specifically determining the presence or absence of a given gene has been developed, so genes have a presence/absence call based on the exons within the same transcript [97].

Cross-hybridization can also be problematic. In the presence of cross-hybridization, an increase in measured intensity will appear, and the measurement may often be higher than constitutively expressed exons. Affymetrix recommends eliminating measurements that appear at a much higher or much lower intensity than other probesets in the gene. Removing highly expressed probesets is not problematic, as the intensity appears higher than other constitutively expressed exons. However, removing poorly expressed exons is problematic, as it is difficult to tell whether poor expression is caused by poor hybridization or by only a fraction of the transcripts in fact containing sequences corresponding to the probeset of interest, in which case the exon positive test is detecting differential alternative splicing.

Another risk is if the probe sequences are annotated incorrectly. Affymetrix

assigns each probeset a "confidence level" based on the line of evidence utilized

to identify the transcript associated with the probes. Selecting to analyze genes

only at the "Core" confidence level reduces the chance of misassignment.

A final potential problem Affymetrix warns about is extremely low variance across

a specific probeset, since the low variance could be mediated by an event such

as the probe not hybridizing to the target sequence. Affymetrix does not

recommend a specific filter for identifying regions with excessively low variance,

but the possibility of an exon showing low variance for reasons independent of

alternative splicing should be taken into account.

Figures

ATP-binding cassette, sub-family G (WHITE), member 5 (Abcg5)

5'   3

| 3' | | | | | | Ace2 - NM_001012006 | | | | | | 5 |

X

Exon Expression

Attribute ▲ Control ■ Treatment   q21
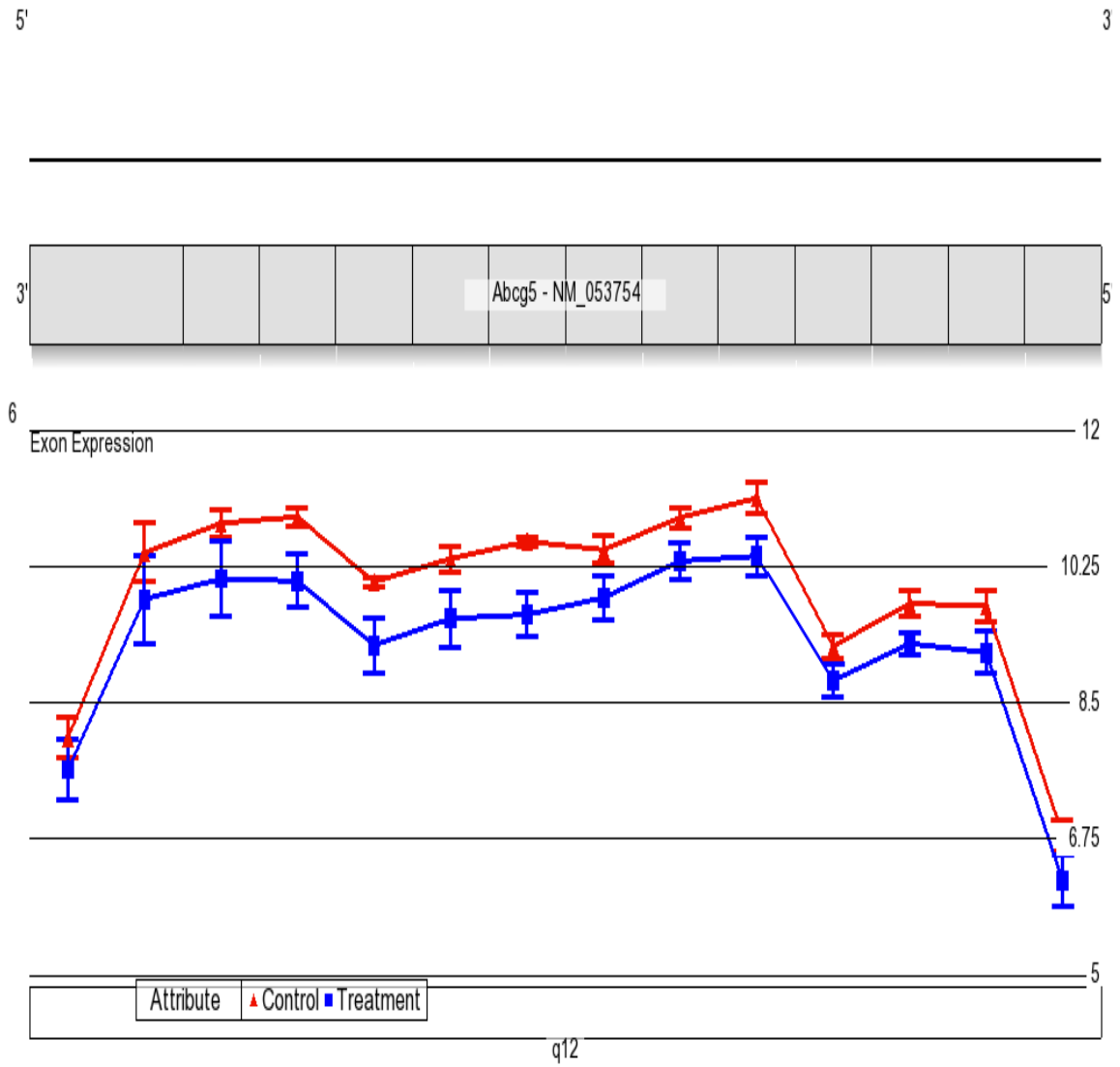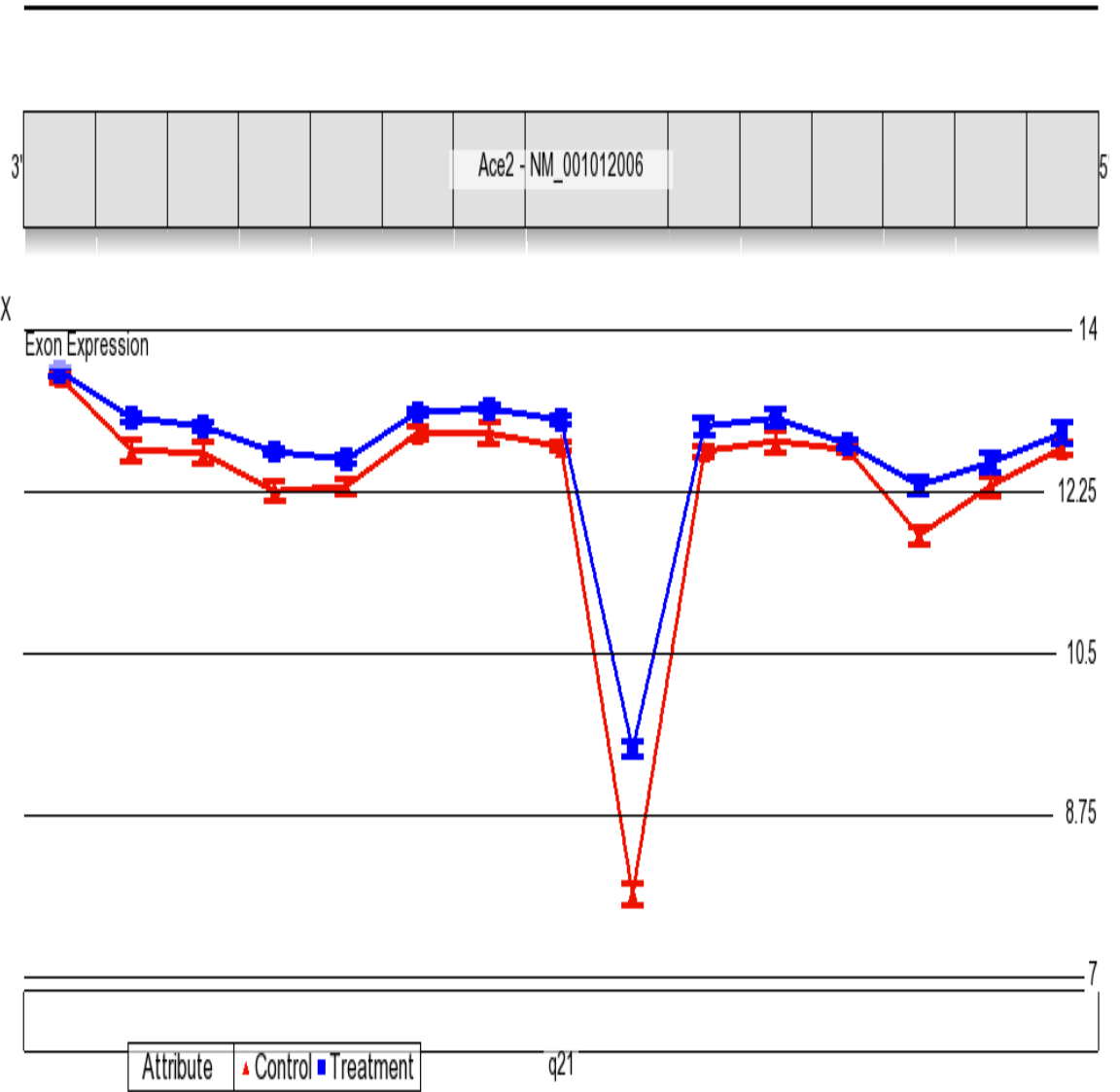
Figure 2.1: Genes that test negative (Abcg5 top) and positive (Angiotensin converting enzyme 2, bottom) for alternative splicing in the ileum. Chromosome position is plotted on the x-axis, chromosome number is listed on the left y-axis, and probeset intensity is plotted on the right y-axis. Note that for the gene (Abcg5) that tests negative, the exon intensities appear to be parallel between control and lactating groups as a function of chromosome position, while this is not true for the gene (Ace2) that tests positive. Along the top of each figure is the gene and its corresponding refseq entry, with the boxes representing individual exons. Both Abcg5 and Ace2 are annotated as being on the minus strand. In order to show both strands, Partek draws genes on the minus strand with the 5' end on the right and the 3' end on the left.

Chapter 3

Differential gene expression in liver and small intestine from lactating rats compared to age-matched virgin controls detects increased mRNA of cholesterol biosynthetic genes

Background

Lactation is a time of a four- to five-fold increased energy demand imposed by the suckling young that requires a proportional adjustment in the ability of the lactating dam to absorb nutrients and to synthesize critical biomolecules to meet the dietary needs of both the offspring and the dam [1-3]. Lactating rats have a two- to three-fold increase in food consumption (hyperphagia) [1-3], in part through the decreased suppression of appetite accompanying decreased serum leptin [19].

Diet and hyperphagia have been shown to influence the rate of cholesterol synthesis, which is increased in the liver and small intestine in the lactating rat [9, 111]. Of these tissues, the liver is the primary contributor to serum levels of cholesterol, and shows a quantitatively greater increase in the rate of cholesterol synthesis during lactation [9]. 3-Hydroxy-3-methylglutaryl-coenzyme A reductase (Hmgcr), the enzyme catalyzing the rate-limiting step of cholesterol synthesis,

shows significantly increased activity in the liver during lactation compared to virgin and nonlactating control rats [86, 112]. Cholesterol synthetic and lipogenic genes are regulated by transcription factors termed sterol regulatory element binding proteins (Srebp); the activity of Srebp proteins is in turn regulated by the Srebf chaperone (Scap) and the Insulin induced genes (Insig).

Circulating serum levels of several hormones that regulate metabolism are decreased during lactation in the rat, including thyroid hormone, insulin and leptin [19]. Such changes in hormone signaling and diet are likely to have large influences on the activation of their corresponding pathways. Receptors for leptin, thyroid hormone, and insulin are expressed in both the liver and small intestine [23, 113-117], with liver being considered a major site of insulin signaling [113] and thyroid receptor β (TRB, Thrb) being the dominant form of the thyroid receptor in both tissues [23]. Leptin acts on the small intestine and inhibits sugar uptake [115], and the liver is a major source of the soluble form of the leptin receptor, particularly under conditions of negative energy balance [116], as occurs in lactation [3]. Therefore, altered serum levels of these hormones would be expected to influence mRNA expression of downstream genes.

Our laboratory has been investigating the effects of lactation on the synthesis and transport of bile acids in the liver and small intestine in the rat [10]. Bile acids are synthesized in the liver from cholesterol, and are essential for the

biliary excretion of cholesterol and for the efficient intestinal absorption of cholesterol, lipid-soluble vitamins and lipids [90]. Bile acids are secreted into bile by the Bile salt export pump (Bsep; Abcb11), taken up across the apical membrane of the enterocyte in the terminal ileum by the Apical sodium-dependent bile acid transporter (Asbt; Slc10a2), effluxed into portal blood by the Organic solute transporter heterodimer (Ostα/β), and then taken up in the hepatocyte by the Sodium-dependent taurocholate co-transporting polypeptide (Ntcp; Slc10a1) [118]. Expression of Ntcp, Bsep and Asbt are all increased in lactation [82, 84], as is the size of the bile acid pool [10]. We recently demonstrated that expression and activity of Cyp7a1, the enzyme catalyzing the rate limiting step in the conversion of cholesterol to bile acids, is increased at mid-lactation (day 10 – 14 postpartum) [10]. Further, this increase occurs at 16 h (10 h of light on a 12 h/light dark cycle; 4 PM) and represents a shift in the diurnal rhythm of *Cyp7a1* expression, which is normally maximal in the dark cycle (i.e., 22 h). Increased expression of *Cyp7a1* is apparently due to decreased expression of Fibroblast Growth Factor 15 (*Fgf15*) in the ileum, resulting in decreased FGF15 signaling via Fibroblast Growth Factor Receptor 4 (Fgfr4) and Erk1/2 in liver and decreased repression of *Cyp7a1* transcription [119].

In order to identify further changes in expression of genes important in the regulation of bile acid and cholesterol synthesis, as well as other genes important in meeting nutritional demands and physiological changes of the lactating rat, we carried out a microarray experiment in the liver and small intestine of the lactating

dam at 16 h on days 10 -11 postpartum and compared these to gene expression

in female virgin control rats.

RESULTS

Detection of differentially expressed genes

A repeated measures mixed model ANOVA was used to test for effects of tissue and lactation, as described in Methods. A statistically significant difference was determined to exist when the physiologic state effect (comparison between all control samples against all lactating samples) yielded a $p< 0.05$ and the effect of lactation within a given tissue (physiologic state simple effect) yielded a $p< 0.01$. Analyzed data are available in Additional File 3.1: Statistical_Analysis_and_Statistical_Pattern_Matching_Results.txt. These p-values were used as cutoffs for differentially expressed genes and led to the proportions of false positives that are listed in the section titled "Approximation of false discoveries" below.

A number (1,114) of genes demonstrated an interaction at $p< 0.01$ and 556 genes passed a Benjamini-Hochberg false discovery rate correction at FDR=0.05 (Additional File 3.2: Benjamini_Hochberg_False_Discovery_Rates.txt). These genes represent those that displayed a different response to lactation in one tissue with respect to the other tissues. However, it should be noted that this list was not a useful cutoff, as genes that responded uniformly to lactation across all tissues would be ignored. Also, five genes that passed the Benjamini-Hochberg correction did not show any significant changes at $p<0.01$ when the effect of lactation was tested within each tissue. Since the primary purpose of this study

was to characterize the influence of lactation on gene expression on these tissues, the interaction term was not used. The tissue effect p-values indicated that many (70% at $p<0.01$) genes were differentially expressed across tissues, due to the large difference in cell types between the liver and the small intestine. Therefore, we chose the cutoffs of a physiologic state main effect at $p<0.05$ and the effect of lactation within at least one tissue at $p<0.01$, as described in the Methods.

Although not as many genes were detected as significantly differentially expressed compared to the overall tissue effect, the overall physiologic state effect and the pairwise comparisons (effect of lactation in each tissue) showed a high number of low p-values, indicating that the tissues in question responded to lactation at the level of mRNA. Several genes were downregulated in the duodenum only (34 genes), with 23 genes showing over a 50% decrease. Members of this group are listed as pattern "-100" in Additional File 3.1. Histograms displaying the distribution of p-values are in Additional File 3.3: Histograms_of_p_values.ppt , and volcano plots displaying each tissue's response to lactation are in Additional File 3.4: Volcano_plots.pdf.

Approximation of false discoveries

The proportion of false positives is an approximation of the ratio of false positives in the list of genes listed as differentially expressed. An estimate of the number

of false positives was calculated for all tissues using genes where p<0.05 for an overall physiologic state effect, and for each tissue using genes with an effect of lactation within each tissue (here defined as a simple effect) of p<0.01. Thus, the proportion of false positives (PFP) was calculated as (p-value cutoff X number of genes tested)/number of genes detected below the p-value cutoff. The PFP as defined by Fernando et al [120], is $E(V)/E(R)$ where $E(V)$ is the expected number of false rejections of the null hypothesis and $E(R)$ is the expected number of rejections of the null hypothesis. Here we utilized the actual number of rejections of the null hypothesis as the expected value. Of all of the genes on the chip, 14,129 genes were found to be annotated and expressed in at least one tissue/ physiologic-state combination and were used for statistical tests. Of those genes, 1,924 had an overall physiologic state p-value of less than 0.05, yielding a PFP of 0.37; at an overall physiologic state p<0.01, 690 genes were detected, yielding a PFP of 0.20. PFPs for the individual tissues at a cutoff of p<0.01 for the pairwise comparisons (simple effects) were 0.17 for the liver, 0.28 for the duodenum, 0.22 for the jejunum, and 0.18 for the ileum. The PFPs for the individual tissue calculations examined only the genes detected at a simple effect p<0.01, and not at the combined overall physiologic state cutoff of p < 0.05 together with the within-tissue cutoff of p< 0.01, as some genes passed the p< 0.01 cutoff within a given tissue, but did not pass the initial overall physiologic state effect cutoff of p< 0.05. The rationale for not utilizing the overall physiologic state effect together with the within tissue physiologic state effect in the calculation was that these multiple tests were utilized for the same gene.

40

Consequently, the list of genes reported at both p<0.05 for the physiologic state

and p<0.01 for the comparison within a tissue was a subset of the list of genes

that only show a p<0.01 within a tissue. The physiologic state cutoff of p< 0.05

was chosen to protect against repeated testing for each tissue; this value was

also chosen because changes that only occurred within one tissue would be

difficult to detect if the overall physiologic state cutoff was made at p< 0.01.

Approximations of the proportion of false positives in this range (0.17 – 0.28)

have been reported previously [121].

RT-PCR Validation of Microarray Data

Results from RT-PCR analyses agreed with the trends detected in the microarray

analyses (Table 3.1). In some cases, significance calls differed, but the

directionality of the changes observed was consistent with the microarray data.

Possible causes for disagreement included the fact that different methods of

normalization were used between RT-PCR and the microarray.

Patterns

Patterns were identified using statistical pattern matching [122, 123] by assigning

each gene as significantly "up", "down", or "no change" detected in each tissue.

The results of the statistical pattern matching showed that fifteen genes were

upregulated in all four tissues, while thirty-one genes were downregulated in all

four tissues. Seventy-two genes were upregulated in liver only, and another ninety-nine genes were downregulated in liver only. Results from analyses of each pattern using DAVID [104] are shown in additional file 3.5: DAVID_output_file.txt.

Of the fifteen genes upregulated in every tissue (Table 3.2), seven were identified by DAVID as being involved in the Biosynthesis of Sterols pathway, where $p=4.59 \times 10^{-13}$, using the list of "Up in All Tissues" for the DAVID analysis. One gene, *transmembrane protein 97* (*Tmem97*), has been identified as being regulated by the Srebp proteins [88, 124], and was recently suggested to aid in Low density lipoprotein receptor (Ldlr) function [124]. Amongst genes not associated with biosynthesis of sterols, *RNA (guanine 9) methyltransferase domain containing 2* (*Rg9mtd2*), is the homolog for a tRNA methyl transferase that occurs in yeast [125]. *Slc39a4_predicted* was upregulated in all tissues, consistent with increased zinc absorption during lactation, and is discussed further below.

Thirty-one genes were identified as downregulated in all tissues (Additional File 3.6: Genes_with_decreased_mrna_all_tissues.pdf). According to the over-representation analysis in DAVID, the KEGG T cell receptor pathway was over-represented in this group (p=0.004) [105, 126] (Additional File 3.5), although only three genes appeared in this list. This pathway did not pass any of the multiple testing procedures available in DAVID, but is consistent with the IPA results,

which flagged "T-cell signaling and differentiation" to be downregulated in the jejunum and ileum (Figure 3.1).

Genes downregulated only in the duodenum were also investigated using DAVID, as several genes revealed a strong downregulation in this tissue (Additional File 3.5). Many of these genes have been identified as being expressed in the pancreas, i.e., eight of the 34 genes in this group matched the Sp_PIR keyword "pancreas" ($p=5.00X10^{-15}$). The function of these genes in the duodenum and the reason for their poor expression in lactation is not known.

The list of genes upregulated in all parts of the small intestine was not significantly enriched by any KEGG pathways in DAVID. The term Lipid Biosynthetic Process was overrepresented ($p=0.001$), although this term did not pass any multiple testing correction available in DAVID.

"The Fibronectin Type III fold" Interpro entry was flagged as overrepresented in the list of genes downregulated in the small intestine, but not in liver. [The genes associated with the Fibronectin Type II fold entry were *Interleukin receptor 22, alpha 2 (Il22ra2)*, *Immunoglobulin superfamily 9* (*Igsf9), Insulin receptor (Insr), Rims binding protein 2 (Rimbp2),* and *Protein tyrosine phosphate receptor type g* (*Ptprg)].* However, the relevance of the downregulation of these genes in lactation is not known*.

Categories overexpressed in the list of genes upregulated in the liver only included the gene ontologies for "response to nutrient levels" and "cholesterol metabolic process". The list of genes downregulated in the liver showed the gene ontology associated with positive regulation of programmed cell death and may partially explain the increased liver size during lactation [8, 127].

Bile acid biosynthesis

Expression of *Cyp7a1*, the enzyme catalyzing the rate limiting step of bile acid biosynthesis, was detected to be increased (p=0.0002) in the liver with a 1.76-fold change. Few other changes were detected in the bile acid biosynthetic pathway. Expression of *Cyp46a1, Ch25h, Cyp27a1, Cyp39a1, Cyp7b1, Cyp8b1, Akr1d1, Slc27a5, Acox2, Scp2*, and *Baat* [128] did not show a significant change in the liver, suggesting that the increase in bile acid biosynthesis observed was triggered by the increase in Cyp7a1 mRNA [10]. These data are consistent with our earlier detailed characterization of mRNA and protein expression of Cyp7a1, Cyp27a1 and Cyp8b1 in lactation [10].

Ingenuity Pathways Analysis

The lists of differentially expressed genes for each tissue, based on the overall physiologic state effect p-value and the respective simple effects were examined by IPA. The three overrepresented pathways with the lowest p-values in each tissue are shown in Table 3.3, and selected pathways are discussed below.

Cholesterol synthesis and metabolism

"Biosynthesis of steroids" had the lowest p-value among IPA's "Canonical Pathways" in three of the four tissues, with the jejunum being the exception. At the designated cutoff (p< 0.01), the jejunum showed a much more modest change in the "biosynthesis of steroids" pathway (p = 0.011 for overrepresentation in jejunum; $p<1 \times 10^{-6}$ in all other tissues). Detailed visualization of the pathway revealed that the upregulated sections of the "biosynthesis of steroids" pathway corresponded with cholesterol synthesis (Additional Files 3.7-3.10: Biosynthesis_of_sterols_in_liver.jpg, Biosynthesis_of_sterols_in_duodenum.jpg, Biosynthesis_of_sterols_in_jejunum.jpg, and Biosynthesis_of_sterols_in_ileum.jpg). Since cholesterol synthesis is regulated by Srebp proteins, Srebp-regulated genes were investigated further. To determine if an exceptionally large number of Srebp-regulated genes were in the list of differentially expressed genes, a list of genes shown to be regulated by Srebp by detection through microarray analysis in Srebp-overexpressing and Scap knockout mice was used for a right-tailed Fisher's Exact test [87] using an online calculator (www.Langsrud.com/Fisher.htm). Here, a p-value of <0.01 in the tissue being tested was defined as a positive test for the purpose of determining whether a given gene was differentially expressed.

45

Of the 33 genes reported to be regulated by nuclear Srebp proteins [87], 29 were present in the data set. A Fisher's exact test p-value of p<0.001 was calculated for each tissue, with the jejunum (p=0.00029) having the fewest genes displaying a significant change (eight genes). Additional File 3.11: Genes_Regulated_by_Srebp_proteins.pdf lists members of the cholesterol biosynthetic pathway and other genes that have been shown to be differentially expressed in Srebp-overexpressing mice and Scap knockout mice [87] and indicates the p-values for each tissue and the ratio of the background-corrected, normalized, untransformed intensities (lactation intensity/control intensity). As shown in Figure 3.2, Hmgcr mRNA expression was increased in three of the four tissues, with no change detected in the jejunum (p = 0.14). mRNA expression for the genes (*Srebf1, Srebf2*) encoding the Srebp proteins were not differentially expressed, although a tendency for a change was detected in the jejunum, where the simple effect comparison p-value for Srebf1 was 0.0006, and the overall physiologic state effect p value was 0.054. Expression of Insig1 mRNA, which is regulated by Srebp activity, showed a significant increase in each part of the small intestine. In contrast, an increase in Scap mRNA occurred in the liver only (p = 2.18E-7). Ldlr mRNA was upregulated in the duodenum and jejunum, and Tmem97 mRNA, an Srebp target [124], was upregulated in all tissues; both Ldlr and Tmem97 proteins aid in LDL uptake by cells [124].

Cholesterol uptake

46

The only gene known to mediate cholesterol uptake in the gut, Npc1l1 [129], is not contained in the extended dataset for the Affymetrix Rat Exon 1.0ST. Investigation of the "full" and "all" datasets indicated that no probeset on the chip was annotated as Npc1l1. Therefore, expression of Npc1l1 was investigated by RT-PCR (Table 3.1). No significant changes were detected in Npc1l1 expression in any tissue. Abcg5 and Abcg8, which function as a heterodimer to efflux cholesterol from the enterocyte into the gut lumen and from the hepatocyte into bile [130], showed decreased expression in the liver (Abcg5 $p=2.9X10^{-7}$; Abcg8 $p=3.6X10^{-6}$) and in the ileum (Abcg5 $p=0.0062$; Abcg8 $p=0.0048$) (Table 3.4).

Transporters

The ATP binding cassette (ABC) transporters that showed a significant change in at least one tissue were also investigated (Table 3.4). The ABC transporters are a superfamily of membrane transporters with diverse substrates that in eukaryotes mediate the ATP-dependent efflux of endogenous substrates, including bile acids and cholesterol, as well as of xenobiotics, including many drugs. The ABC transporter Abcb1a (Mdr1a) was downregulated in every tissue. This protein effluxes xenobiotics across the apical domain of the hepatocyte into bile, and in the enterocyte and plays an important role in limiting absorption of orally administered substrates [131]. Two members of the Abcc (Mrp) subfamily,

Abcc5 and Abcc6, were downregulated in the ileum, while Abcg2 expression was increased in the duodenum.

Solute carrier proteins (Slcs) are a superfamily of proteins that transport many different molecules, including amino acids and ions (www.bioparadigms.org). All Slcs on the chip were investigated and those with a detected significant change in lactation in any tissue are shown in Additional File 3.12: Slcs.pdf. For each tissue, nearly 20% of the Slcs showed a change at an overall physiologic state effect of $p < 0.05$, and approximately 5% were declared significant after applying a $p < 0.01$ cutoff within a tissue (simple effect)

Slc39a4 is a transporter mediating the uptake of zinc into the intestine [132]. *Slc39a4_predicted* was one of fifteen genes to be significantly upregulated in every tissue ($p < 1 \times 10^{-5}$) (Additional File 3.1). Fold changes for *Slc39a4 predicted* based on untransformed intensity values ranged from 1.65 in the jejunum to 2.81 in the liver (Table 3.1).

Thyroid signaling

A Fisher's exact test using IPA detected significant overrepresentation in the TR/RXR pathway in every tissue. (Additional Files 3.13-3.16: Liver_canonical_pathways.txt, Duodenum_canonical_pathways.txt., Jejunum_canonical_pathways.txt., and Ileum_canonical_pathways.txt). Thyroid

hormone receptor α (TRA, Thra) and thyroid hormone receptor β (TRB, Thrb) were both downregulated in the ileum and jejunum. A decrease was also seen in the liver, but the change was not significant (p = 0.07 for TRA; p = 0.018 for TRB). Surprisingly few of the downstream genes of TR/RXR were downregulated in the IPA depiction of this pathway. In some cases, overlap occurred with Srebp signaling, and increased signaling from Srebp appeared to have overridden decreased thyroid signaling. This seems to have occurred with *Acetyl-CoA Carboxylase alpha (Acaca)* in the liver and *Ldlr* in both the duodenum and jejunum [87]. One TRB/RXR regulated gene [24], *Kruppel like factor 9 (Klf9)*, which is a transcription factor associated with intestinal proliferation, was downregulated in every tissue (Additional Files 3.17-3.20: Liver_thyroid_pathway.jpg, Duodenum_thyroid_pathway.jpg, Jejunum_thyroid_pathway.jpg, and Ileum_thyroid_pathway.jpg). Klf9 knockout mice have shorter intestinal villi, although Klf9 is typically considered a transcriptional repressor and can also negatively regulate growth [33].

Decreased mRNA from T-Cell receptor signaling and related pathways

mRNA of genes coding for the components of T-Cell receptor signaling pathway in IPA showed significant downregulation in the jejunum and ileum (p = 1.68×10^{-7} and p = 4.7×10^{-4}, respectively) (Figure 3.1). A similar pathway, the "CD28 receptor signaling in T helper cells" pathway was also downregulated in the jejunum, but substantial overlap between the two pathways suggested

observation of the same events. These pathways are upstream of IL-2

production [133], however, the microarray detected no change in IL-2 mRNA in

any tissue.

Discussion

Cholesterol Biosynthesis

IPA and DAVID both flagged "Biosynthesis of Steroids" to be overrepresented in the list of differentially expressed genes in lactation in three of the four tissues (the jejunum displayed a p-value near the cutoff), and the list of genes upregulated in all tissues during lactation, respectively (Additional Files 3.7-3.10). Visualization of this pathway revealed that the genes identified were components of the cholesterol biosynthetic pathway. Statistical pattern matching as well as results from overrepresentation analyses in both DAVID and IPA indicated that expression of cholesterol biosynthetic genes was induced in all tissues examined, although to a lesser extent in the jejunum (p=0.011; Benjamini-Hochberg corrected p=0.048, values calculated by IPA). Cholesterol and lipid biosynthetic genes are known to be regulated by transcription factors known as the Srebp proteins. Three Srebp proteins are encoded by two genes, *Srebf1* and *Srebf2*. *Srebf1* codes for Srebp-1a and Srebp-1c, while *Srebf2* codes for Srebp-2. The Srebp proteins differ in their control of fatty acid synthesis and cholesterol biosynthesis [88, 91, 134]. Srebp-2 is associated with cholesterol biosynthesis, while the Srebp1 proteins are associated with fatty acid synthesis [88], although there appears to be overlap in the genes that are responsive to these transcription factors [87]. Srebp-1c is sensitive at the transcriptional level to LXR signaling [91, 135, 136], while all three share a mechanism for becoming an

active transcription factor [91]. Under conditions of sufficient cholesterol concentrations, the protein product of Insulin stimulated gene (Insig) binds to Srebf Chaperone (Scap) to retain an Insig/Scap/Srebp complex in the endoplasmic reticulum [91, 137]. In the absence of oxysterols and cholesterol, Insig is degraded and Scap is released from the endoplasmic reticulum; Scap then escorts the bound Srebp to the Golgi, where the N-terminus of the Srebp is cleaved from the full protein to generate the active form that functions as a transcription factor [134]. As indicated above, cholesterol synthesis is increased during lactation in both the liver and the small intestine [2, 3] and Srebp target genes have also been shown to be upregulated in mammary tissue in lactating dams [89]. In a review article, Shimano lists in a table (Shimano, Table 1) a series of genes with known SRE elements and their promoter sequences [138]. Of the genes in this table, low density lipoprotein receptor (Ldlr), HMG CoA synthase1 (Hmgcs1), HMG CoA reductase (Hmgcr), farnesyl diphosphate farnesyl transferase 1 (Fdft1), Cytochrome P450 family 51 (Cyp51), Acetyl CoA carboxylase (Acc), Stearoyl CoA desaturase 2 (Scd2), and malic enzyme 1 (ME1) are all upregulated in at least one tissue.

The observation that Srebp-regulated genes were upregulated is consistent with early data showing an overall increased cholesterol biosynthesis in lactation [9]. We currently do not know which Srebp isoforms are involved in the changes seen in lactation, as Srebp-1c, Srebp-1a and Srebp-2, are all able to regulate expression of cholesterol synthetic genes. However, Srebp-2 plays a stronger

52

role in regulating these genes [88]. The only potential change detected in the mRNA for a Srebf gene was *Srebf1*, the gene associated with Srebp-1a and Srebp-1c, which showed increased mRNA levels in the jejunum.

In the liver, Scap mRNA showed a significant increase ($p = 2.2 \times 10^{-7}$). If this change were associated with an increase in Scap protein levels, then a probable mechanism for the increase in Srebp target genes in the liver would be increased transport of Srebp proteins to the Golgi, and their subsequent delivery to the nucleus [91]. Insig1 mRNA also showed a significant increase in all parts of the small intestine, but not in the liver. Insig1 functions to retain Srebp in the endoplasmic reticulum. Thus, increased expression of Scap in liver and increased expression of Insig1 in intestine provide a likely mechanism for the greater increase in cholesterol synthesis in liver vs. intestine observed by Feingold et al [9].

A number of factors contribute to the increased need for cholesterol in the lactating dam. The dam requires significant cholesterol for the increased synthesis of bile acids; 50% of cholesterol catabolized in the liver from nonlactating rats is used for bile acid synthesis [139]. Since the size of the bile acid pool increases 2-3-fold at 10 -14 d of lactation [10], greater than 50% of cholesterol is likely catabolized to bile acids in lactation. The proportion of dietary cholesterol vs. endogenously synthesized cholesterol that is catabolized to bile acids in nonlactating vs. lactating rats is not known. Most importantly,

cholesterol is an essential component of milk that supports membrane synthesis and neurodevelopment in the pups [140]. About 16 mg per day of cholesterol is secreted into the milk in rats [141]; between 32 and 40% of this cholesterol is synthesized in the mammary gland, while 11% is absorbed from the diet [86]. Thus, cholesterol synthesized in the liver makes up about 50% of cholesterol secreted in milk [86].

In addition to detecting a change in cholesterol synthesis, a possible mechanism for improved net cholesterol uptake was found. *Abcg5* and *Abcg8* show decreased levels of mRNA expression in the liver and ileum. A decrease in the concentration of active Abcg5/Abcg8 heterodimer in the intestine would be expected to yield an increase in net cholesterol uptake through decreased efflux from the enterocyte into the gut lumen, while decreased hepatic expression would minimize cholesterol secretion into bile [93]. The decreased expression of Abcg5/g8 mRNA in the liver, together with increased expression of cholesterol synthetic genes, likely serve to enhance conservation of cholesterol to allow for sufficient transfer of cholesterol into the milk and for synthesis of bile acids. Increased synthesis of bile acids would in turn serve to increase cholesterol absorption [90]. Taken together, these data suggest a concerted mechanism for enhancing net cholesterol absorption and minimizing its elimination to ensure sufficient cholesterol for incorporation into milk and bile acid synthesis, both important factors in maintaining the health of both the dam and pups.

Zinc

Statistical pattern matching found that Slc39a4_predicted mRNA was increased (1.65 to 2.81 fold) in all tissues. Slc39a4 is a major zinc transporter associated with zinc import into the enterocyte [132], and zinc absorption is up-regulated in lactation [73-76]. Taken together, these data imply that the increased expression of Slc39a4 mediates the increased zinc absorption that occurs in lactation. Zinc is an essential nutrient shown to be important in bone development [75], to play a role in stimulating the insulin pathway [142] and in controlling T-Cell activity [143]. Zinc requirements are increased during lactation relative to pregnancy, and therefore net zinc uptake needs to be increased to maintain zinc homeostasis during lactation in humans, particularly during early lactation [144]. Interestingly, alpha-2-macroglobulin (A2m) showed a substantial increase in mRNA expression (~15-fold) in the liver. Zinc can directly regulate A2m's ability to sequester cytokines [143, 145, 146] by enhancing formation of a form of A2m that contains free sulfhydryl groups, which serve as binding sites for the cytokines [146].

Downregulation of mRNA from the T-Cell signaling pathway

Both the jejunum and ileum showed strong downregulation at the mRNA level of the proteins composing the T-Cell signaling pathway in IPA. These changes may reflect a decrease in the number of actual T-Cells in the small intestine of lactating rats.

Xenobiotic Transporters

Table 3.4 displays significantly different changes in mRNA concentration for the ABC transporters. Included in this list of genes are *Abcb1a (Mdr1a), Abcc5, Abcc6, and Abcg2. Abcb1a* showed decreased expression in all tissues in the microarray and this was successfully validated for the jejunum and ileum by RT-PCR. *Abcc5* and *Abcc6* also showed decreased mRNA expression in the ileum. Decreased expression of these efflux transporters would in general lead to an increased net absorption of their substrates. In contrast, Abcg2 showed increased expression in the duodenum, which would decrease absorption of substrates. Sample substrates for these proteins include drugs such as digoxin and cyclosporine A (Abcb1a, [147]), cGMP (Abcc5 [148]), the glutathione conjugate leukotriene C_4 (Abcc6 [149] [150]), and 2-amino-1-methyl-6-phenylimidazo[4,5-b] pyridine, a dietary carcinogen (Abcg2 [151]). Further work is needed to understand the impact of these specific changes within the context of lactation on the lactating dam and her pups.

TR/RXR Pathway

IPA found members of the TR/RXR pathway to be overrepresented in the list of differentially expressed genes in each tissue. In both the ileum and jejunum, both TRA and TRB showed down-regulation. Although lactating rats are hypothyroid

[19] and expressed lower levels of mRNA for thyroid receptor in these tissues, not all thyroid responsive genes were down-regulated, including Apoa5, Eno1, and Glut1, which showed no change in any tissue. A more detailed picture of the thyroid receptor pathway can be found in Additional Files 3.17-3.20. Serum thyroid hormone levels and early steps in the pathway are likely downregulated as an attempt to conserve energy [19, 152]. Fisher's exact test determines its p-values based on the counts of the number of genes in the list of differentially expressed genes and compares these to the total number of genes in the pathway relative to the total number of genes in the microarray. Therefore, pathways that overlap are likely to be detected as overrepresented if the overlapping genes are in the list of differentially expressed genes. Because some Srebp regulated genes are considered to be part of the TR/RXR pathway, the p-values for overrepresentation may be low, even if thyroid signaling overall was unchanged. However, changes in the mRNA levels of the thyroid receptors argue against this, since these receptors have not been shown to be Srebp targets.

Conclusions

The present studies have shown an increase in the mRNA of enzymes involved in the cholesterol biosynthesis pathway, implying that the sterol regulatory element binding proteins are more active in the liver and small intestine in lactating vs nonlactating rats. The data are consistent with a coordinated response to the overall increased energy demands of lactation and the specific needs of the pups for cholesterol so that there is adequate cholesterol for incorporation into milk and increased synthesis of bile acids; the latter in turn function to increase the intestinal absorption of cholesterol and lipids. We also demonstrated a marked increase in the expression of a key transporter important in the uptake of the essential element, zinc. Finally, we detected decreased mRNA from genes associated with T-cell signaling in the jejunum and ileum.

Methods

Animals

Sixteen Sprague-Dawley rats that were lactating for 10-11 days and sixteen age-matched virgin controls were obtained from Harlan (Indianapolis, IN) and maintained on a 12 h light/dark cycle (6 AM lights on/6 PM lights off). Rats had free access to Teklad Global Diet 2018 (Harlan Laboratories, Cincinnati, OH) and water. In order to minimize the variance in energy demands on the lactating dam, pups were removed from large litters within 24 hours of birth so that all litters contained 8-11 pups. All animals were sacrificed at 16 h (10 h of light on a 12 h light/dark cycle; 4 PM), and the liver, duodenum, jejunum, and ileum were removed for total RNA extraction from each tissue. The first 5 cm of the small intestine following the pyloric sphincter was taken as the duodenum, while the 10 cm following the ligament of Trietz was discarded and the next 20 cm used as the jejunum. The 20 cm segment preceding the cecal valve was taken as ileum. The mucosal layer was removed by scraping at 4°C, and used for isolation of RNA from intestinal segments. RNA was extracted from homogenized tissue using Trizol (Invitrogen, Carlsbad, CA), and purified using RNeasy Mini Kit DNAse and columns (Qiagen, Valencia, CA). The integrity of all RNA samples was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Animal protocols were conducted in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals and

59

were approved by the Institutional Animal Care and Use Committee of the University of Kentucky.

Each rat was assigned to one of four pools within the respective physiologic state (four control pools and four lactating pools). Pooled RNA samples, consisting of the RNA from the four rats within the same group, were created for each tissue, with individual rats composing the pools consistent across tissues. Each pooled sample (RNA from one tissue from one set of four rats) was loaded onto a separate chip. This resulted in the use of 32 chips (4 tissues X 2 "physiologic states" X 4 pools). Samples were prepared and processed according to the manufacturer's instructions by the University of Kentucky Microarray Core Facility (Lexington, KY).

Selection of Genes on Which to Perform Statistical Analysis

Affymetrix Expression Console software was used to perform the Robust Multichip Average (RMA) [99, 100] algorithm, which background corrected, quantile normalized, and $\log_2$-transformed gene level summaries of the Extended dataset. Affymetrix has divided the chip into various datasets, which represent different confidence levels with respect to the complementarities between the probe and the sense strand of the gene sequences. The Extended dataset consists of the Core dataset, which is made up of Refseq entries and full length mRNAs, as well as additional multiple annotations based on cDNA libraries.

Although the Core dataset probes are the best annotated, the rat genome is not as well annotated as the human and mouse genomes, with many of the genes in the rat Extended dataset identified based on their similarity to human or mouse genes. Use of the Extended dataset allowed a more thorough analysis of the genome. The summarized values are an average taken across all exons.

The exon level Affymetrix DABG (Detection Above Background) values and the Affymetrix annotation file (version raex_1_0-st-v1.na27.rn4) were used to filter the data [110]. Exon level data was opened in Expression Console and the RMA algorithm and $\log_2$ transformations were performed. An exon was considered present if it had a DABG p < 0.01, indicating that the exon in question had an intensity greater than 99% of the background probes with the same GC content [110]. A gene was considered for analysis if at least one exon was detected on at least two chips within the same tissue and the same physiologic state, e.g., presence on two control liver chips. Genes were also removed if the Affymetrix annotation file contained a "---" or a blank for the "mRNA description" entry (Annotation file: raex_1_0-st-v1.na27.rn4). Of the 19,434 genes in the Extended dataset, 14,129 were utilized for statistical tests based on these criteria.

Statistical Analysis

Since all four tissues were taken from each rat, a repeated measures mixed model ANOVA was used to determine if changes in expression were statistically

significant for each gene [153]. JMP genomics version 4.1 (SAS Institute Inc, Cary, NC, 2011) was used to perform the ANOVA using compound symmetry to model the covariance matrices. Tissue, physiologic state, and the tissue*physiologic state interaction were treated as fixed effects, while "pool", a variable describing the combination of four individual rats to create a sample, was treated as a random subject variable.

A common method for addressing multiple testing issues involved in the analysis of microarrays is the Benjamini-Hochberg false discovery rate (FDR) correction. Several possibilities existed for attempting to address the issue of multiple testing. First, the physiologic state p-value could be adjusted to a fixed FDR. Alternatively, the individual unadjusted p-values for the simple effects could be set to a given value and an approximation of the number of false positives within a group could then be calculated. We chose the latter approach, set $p < 0.01$ as a cutoff and calculated the proportion of false positives (PFPs) for the simple effects that represented the pairwise comparisons within a tissue. We chose this method to balance the risk of false positives with the risk of false negatives. False negatives in the list of differentially expressed genes could interfere with downstream pathway analyses.

The overall physiologic state effect of $p < 0.05$ also served as an additional cutoff to reduce the total number of statistical tests performed. Selection of an overall physiologic state p-value of $p < 0.01$ was problematic, as genes that displayed

differential expression in only one tissue might not be noticed due to the lack of change in the remaining tissues. In summary, genes were considered differentially expressed between control and lactation within a tissue if a significant physiologic state effect was observed at $p<0.05$ and a simple effect for the pairwise comparison within a tissue was $p< 0.01$.

.CEL files and .CHP files describing the data are available through GEO under GSE19175 and analyzed data is available in Additional File 3.1.

Detection of Biological Trends

Ingenuity Pathways (IPA) (www.ingenuity.com) was used to screen the results for biological trends. Differentially expressed genes were determined as described in the "Statistical Analysis" section above. The Rat Exon 1.0 ST chip was used as a background list.

A right tailed Fisher's Exact test was used to screen the IPA database and detect categories that were overrepresented based on genes detected to have a significant difference in expression. Available in IPA is the Benjamini-Hochberg correction, which attempts to control the number of false positives. This calculation is dependent on the size of the database.   We chose not to use the Benjamini-Hochberg correction, but chose to fix the significance threshold at $p<0.01$ for any given test. A $p< 0.01$ for the Fisher's Exact test indicated that

more genes in the list of differentially expressed genes appeared in a pathway than would be expected by chance if the same number of differentially expressed genes were to be selected randomly from all genes on the chip. Both these p-values and the Benjamini-Hochberg p-values are provided in Table 3.3.

The lists generated by statistical pattern matching were analyzed by DAVID [104, 105] with the 14,129 genes used for statistical analyses as the background list. The lists screened for overrepresentation were the Canonical Pathways category in IPA and several databases in DAVID (see below). For the Srebp transcription factors, a report of the differential gene expression in mice overexpressing isoforms of Srebp and in Scap knockout mice was used to create an additional list of genes known to be regulated by the Srebp proteins [87]. Out of the 33 genes listed as regulated by nuclear Srebp-1a and Srebp2, 29 were identified as being on the chip at the level of the Extended dataset and were used to perform a right tailed Fisher's Exact test.

We used IPA's Canonical Pathways database for all changes detected within a tissue, while we used all Gene Ontology terms, COG ontology, Sp_PIR Keyword [154, 155], UP_SEQ_Feature [154, 155], Interpro [156], PIR_Superfamily [157], SMART [158, 159], and KEGG [126, 160, 161] as databases in DAVID for testing each pattern detected by statistical pattern matching.

Statistical Pattern Matching

In order to assign changes of RNA expression into biologically meaningful groups, a method of statistical pattern matching was used analogous to the one used by Arzuaga et al [122] and Hulshizer and Blalock [123]. mRNA from any given gene could increase expression, decrease expression, or show no change in expression in samples from lactating animals compared to controls in each tissue. The method for pattern matching is described in a step by step manner as follows. Only genes that tested positive for differential expression were used for pattern matching. These genes were assigned an additional significance call at $p<0.05$ for each simple effect within a tissue. This was done to protect against incorrectly assigning a gene that systematically had low p-values as differentially expressed in only one tissue. For example, if a gene was downregulated with p values of less than 0.01 in the liver, duodenum, and jejunum, and a p of 0.03 in the ileum, the gene was assigned as differentially expressed in all tissues rather than in three of the four tissues. Although the gene would not have been considered differentially expressed for the purposes of evaluating which genes were differentially expressed in the ileum, the gene was considered differentially expressed for the purposes of assigning a pattern. This reduced the risk of falsely assigning the gene to another pattern in the presence of a false negative. The fold-change within a given tissue was then utilized to determine if a gene was upregulated or downregulated in each tissue. For each tissue, three values were multiplied to define the change in that tissue. One value was an integer that represented the tissue itself: 1000 for liver, 100 for duodenum, 10 for

65

jejunum, and 1 for ileum. The second value was a multiplier to define whether or not there was a tendency (p<0.05) for the gene to change expression in the given tissue, where p<0.05 = 1; p>0.05 = 0). The third value defined the directionality of the change, where upregulation = 1 and downregulation = - 1. The products of these values were used to define the change that occurred within a given tissue. For example, a gene upregulated in the liver would have a value of 1000, i.e., 1000*1*1. The sum of this value for all tissues was taken to generate a unique value for every set of possible changes that could occur across tissues. For example a gene that was upregulated in every tissue except the ileum, where it was downregulated would have a value of 1009 (1000 + 100 + 10 - 1). The following groups were defined to be of particular importance and were investigated: up in all tissues, down in all tissues, up in all parts of small intestine, down in all parts of small intestine, up in liver, down in liver, and down in duodenum only.

RT-PCR validation of microarray findings

The same total RNA samples used for the microarray were used for RT-PCR validation; RT-PCR was performed as reported previously [162], using the geometric average of Ctsb, Tmbim6, and Tmed2 as a normalization constant [10]. An aliquot of all cDNA samples was used to generate a standard curve. RT-PCR experiments were performed by Dr. Tianyong Zhao in our laboratory.

66

Choosing an appropriate normalization constant was performed using methodology similar to that described by Andersen et al [163]. Methods for detecting valid internal controls require a list of candidate genes with little or no bias, since any bias that exists within the genes detected as valid internal controls would be transferred over to the newly calculated normalization factor [163]. Genes with high expression (mean log_2 intensity from the microarray greater than or equal to 10) were sorted by their coefficients of variance. Microarray data for the twenty genes with the lowest coefficients of variance were input into Normfinder [163]. The single best gene to use as a control was Ctsb, and the best combination of two was found to be Tmbim6 and Tmed2. We performed RT-PCR on these genes and chose to use the geometric mean of the three genes as a normalization constant. The same statistical model that was used to analyze the microarray results was used to analyze the normalized RT-PCR data.

cDNA was synthesized using High-Capacity cDNA Reverse Transcription Kits from Applied Biosystems (Foster city, CA) according to the manufacturer's instructions. Primers and Universal Probes Library (UPL) probes for real time RT PCR were designed and ordered from Roche Applied Science (Mannheim, Germany) using online software (www.universalprobelibrary.com). A list of primers is provided in Additional File 3.21: RT_PCR_primers.pdf. The Roche Light Cycler 480 was used for performing the RT-PCR. Briefly, 1 µg of total RNA was used for cDNA synthesis, the synthesized cDNA diluted to 500 µL, and 5 µL

of diluted cDNA used as template in a 20 µL reaction volume.   For quantification

analysis of real time RT-PCR data, a standard curve was generated by pooling

all the cDNA samples to form one cDNA mixture and then diluting this cDNA

mixture 10-, 100-, 1000- and 10,000-fold so that expression of the gene of

interest was within the range of the standard curve.

Figures

T Cell Receptor Signaling
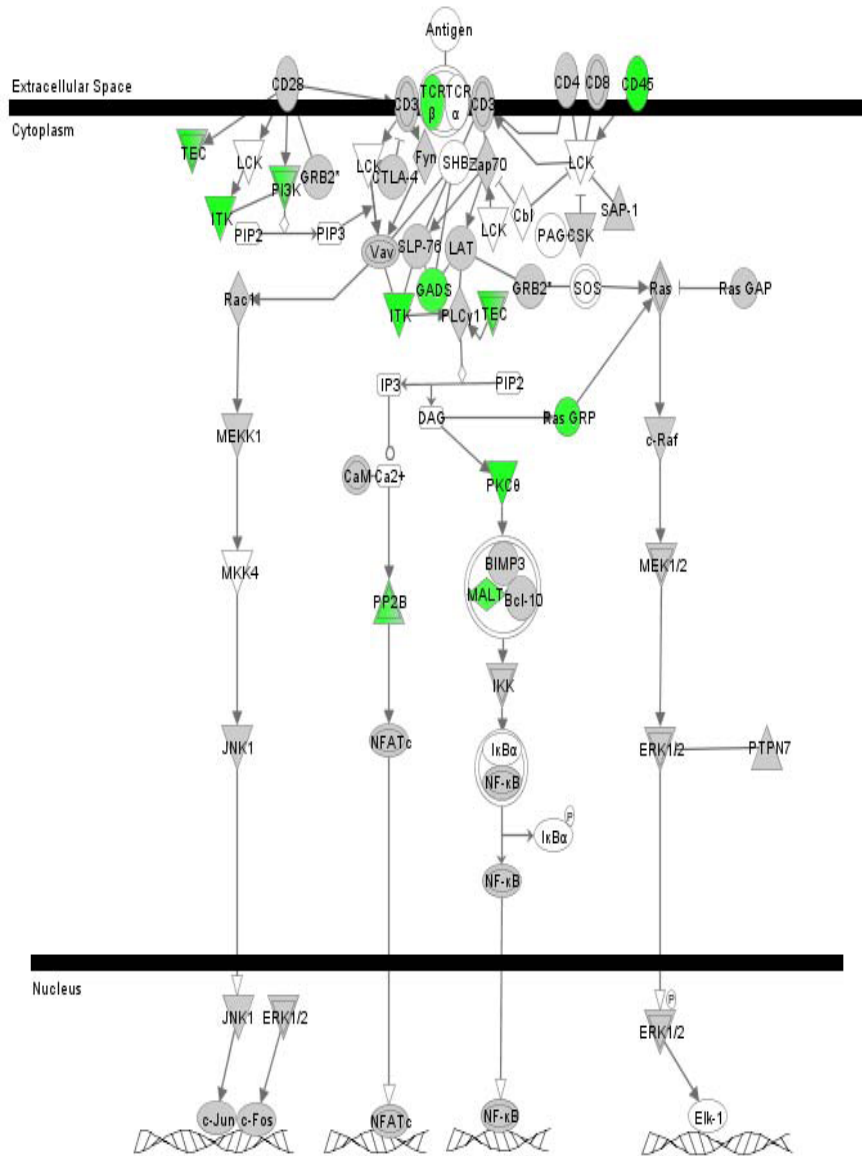
69

T Cell Receptor Signaling
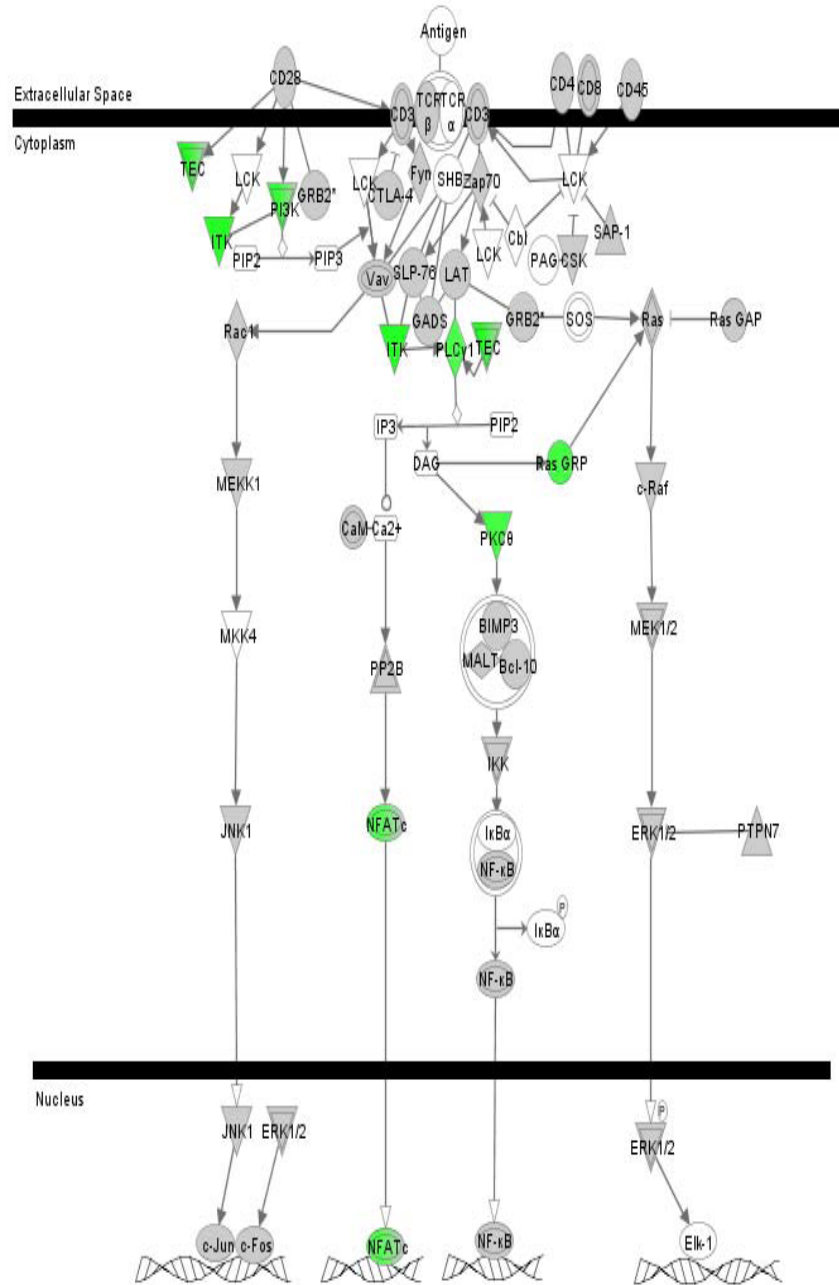
Ileum

Figure 3.1. T-cell signaling in lactating jejunum and ileum

Downregulation of T-cell signaling in the jejunum (top, $p = 1.68 \times 10^{-7}$) and ileum

(bottom, $p = 4.7 \times 10^{-4}$). Green shading indicates downregulation of the

corresponding gene during lactation.
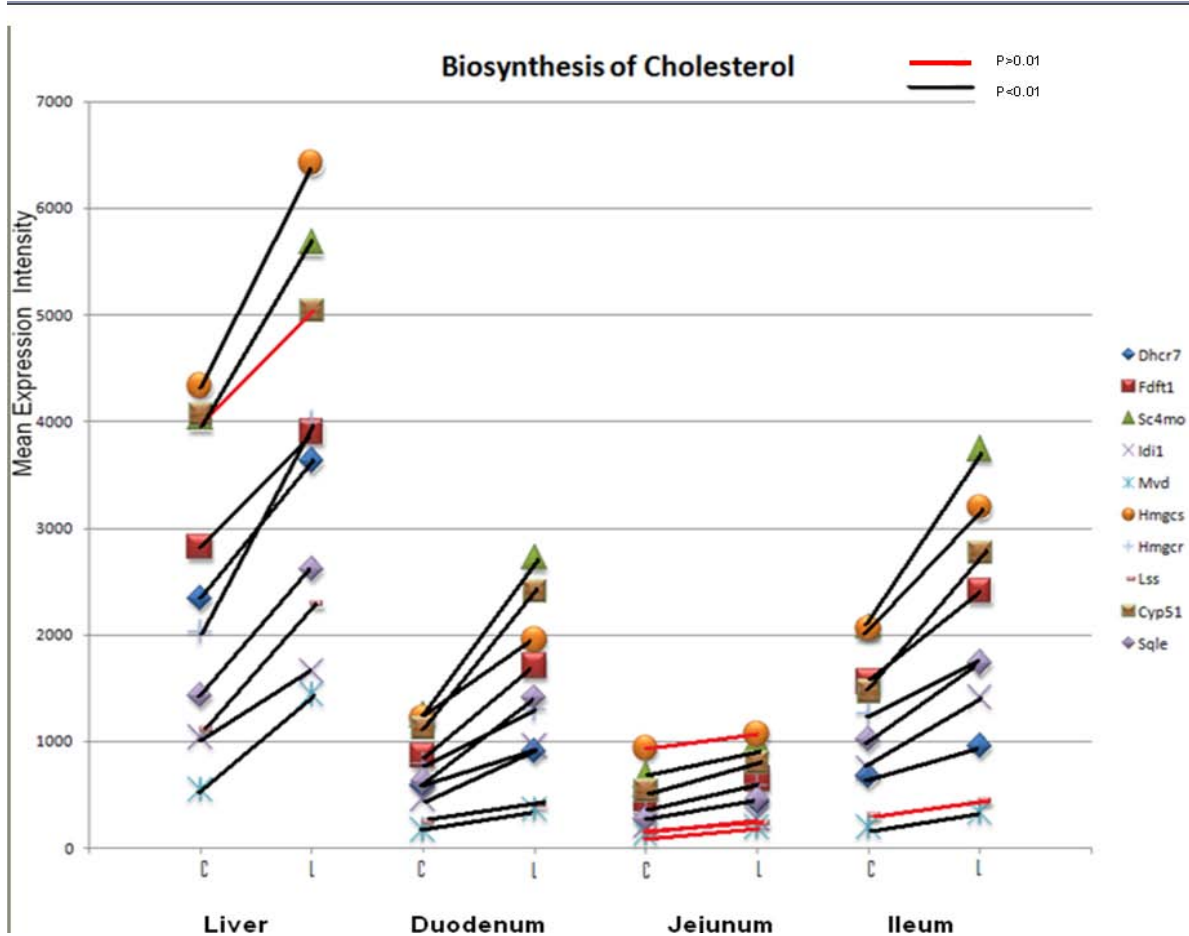
Figure 3.2. Plot of Mean Intensities for Known Cholesterol Biosynthetic Enzymes

The mean untransformed intensities were plotted against tissue/physiologic state combination. C and L indicate the mean of Control and Lactating samples, respectively, within each tissue. Abbreviations for genes are as follows. Dhcr7: 7-dehydrocholesterol reductase; Fdft1: farnesyl-diphosphate farnesyltransferase 1; Sc4mol: sterol-C4-methyl oxidase-like; Idi1: isopentenyl-diphosphate delta isomerase 1; Mvd: mevalonate (diphospho) decarboxylase; Hmgcs: 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1; Hmgcr: HMG Coenzyme A reductase;

72

Lss: lanosterol synthase; Cyp51: cytochrome P450 family 51; Sqle: Squalene

Epoxidase. A red line between Control and Lactating symbols for a given gene

indicates that p>.0.01 for the tissue pairwise comparison within the tissue from

the mixed models repeated measures ANOVA on the $\log_2$ transformed

intensities; the change was not considered significant for the purposes of

determining a list of differentially expressed genes. A black line between C and L

symbols indicates p< 0.01 as described above.

Table 3.1. RT-PCR validation of selected genes from the microarray.

Gene Symbol	Transcript Cluster ID	RL	RD	RJ	RIL
	7080099	1.52	2.60	1.59	1.73
		(p=0.021)	(p<0.01)	(p=0.42)	(p=0.072)
		1.92	2.03	1.82	2.15
Tmem97		(p<0.01)	(p<0.01)	(p<0.01)	(p<0.01)
	-	1.00	1.27	1.22	0.88
Npc1l1		(p=1.00)	(p=0.49)	(p=0.16)	(p=0.72)
	7250393	0.82	0.83	0.72	0.65
		(p=0.91)	(p=0.77)	(p=0.09)	(p<0.01)
		0.63	0.66	0.65	0.70
Abcb1a		(p<0.01)	(p<0.01)	(p<0.01)	(p=0.01)
	7336681	2.98	1.12	2.06	0.78
		(p=0.93)	(p=0.64)	(p<0.01)	(p=0.89)
		2.29	1.23	1.71	0.67
Cyp1a1		(p<0.01)	(p=0.35)	(p=0.02)	(p=0.03)
	7100149	1.83	1.30		
		(p=0.12)	(p=1.00)	BDL	BDL
		1.69	1.23	1.05	1.04
Cyp3a23/3a1		(p<0.01)	(p=0.35)	(p=0.50)	(p=0.58)
	7139070	1.36	2.58	1.02	2.41
		(p=0.11)	(p=0.22)	(p=0.99)	(p=0.018)
		1.38	1.98	1.42	1.55
Fdft1		(p<0.01)	(p<0.01)	(p<0.01)	(p<0.01)

	L	D	J	IL
7202670	2.05	1.52	1.14	1.18
	(p<0.01)	(p=0.30)	(p=0.80)	(p=0.48)
	1.97	1.64	1.17	1.38
Hmgcr	(p<0.01)	(p<0.01)	(p=0.14)	(p<0.01)
7329323	2.71	2.92	2.20	1.99
	(p=0.59)	(p<0.01)	(p<0.01)	(p<0.01)
	2.81	1.89	1.65	1.68
Slc39a4	(p<0.01)	(p<0.01)	(p<0.01)	(p<0.01)
7317317		2.66		
	1.80	(p=0.034)	1.71	1.59
	(p<0.01)		(p=0.50)	(p=0.056)
	1.83	2.27	1.72	1.70
Sqle	(p<0.01)	(p<0.01)	(p<0.01)	(p<0.01)
7117373	0.85	0.76	0.39	0.62
	(p=0.047)	(=0.41)	(p=0.21)	(p=1.00)
Ugt2b36	0.96	0.57	0.37	1.00
(Ugt2b4)	(p=0.86)	(p<0.01)	(p<0.01)	(p=0.97)

Ratios were calculated as Lactating measurement divided by Control measurement. RT-PCR data is reported as the ratio of normalized measurements, and microarray data is reported as the ratio of normalized, untransformed intensities. p-Values were calculated using a mixed models approach as described in methods. RL, RD, RJ and RIL are the Ratio of Lactation to Control in liver (L), duodenum (D), jejunum (J), and ileum (IL), respectively. BDL, below detection limit. RT-PCR results are listed as the top set of values, and microarray results are listed below.

Table 3.2. Genes that displayed increased mRNA during lactation in all tissues

Transcript Cluster ID	Gene Name	RL	RD	RJ	RIL
7027017	Acetyl-Coenzyme A acetyltransferase	1.98 (p=0.0173)	2.40 (p=0.0049)	1.80 (p=0.0363)	2.19 (p=0.0128)
7080099	Transmembrane protein 96	1.92 (p=0.0001)	2.03 (p=2.8e-05)	1.82 (p=0.0002)	2.15 (p=8.7e-06)
7113785	Hydroxysteroid(17-beta) dehydrogenase 7	1.71 (p=2.1e-06)	1.54 (p=4.7e-05)	1.19 (p=0.045)	1.51 (p=0.0001)
7132836	PDZ binding kinase predicted*	1.56 (p=0.0222)	1.88 (p=0.0031)	1.66 (p=0.0054)	1.54 (p=0.0227)
7139070	Farnesyl diphosphate farnesyl transferase	1.38 (p=0.0009)	1.98 (p=4.7e-08)	1.42 (p=0.0004)	1.55 (p=4.1e-05)
7144691	Sterol-C4-methyl oxidase-like	1.41 (p=0.0015)	2.19 (p=1.9e-07)	1.45 (p=0.0008)	1.80 (p=8.2e-06)
7166170	Isopentenyl-diphosphate delta isomerase	1.61 (p=0.0006)	2.11 (p=2.8e-06)	1.33 (p=0.0226)	1.91 (p=2.6e-05)
7169182	Kinesin family member 20a_predicted*	1.46 (p=0.0012)	1.50 (p=0.0010)	1.33 (p=0.0104)	1.45 (p=0.0013)
7186293	Mevalonate (diphospho) decarboxylate	2.65 (p=3.8e-05)	2.15 (p=0.0004)	1.49 (p=0.0316)	1.73 (p=0.0042)
7199743	RNA (guanine-9-) methyltransferase domain containing 2	1.18 (p=0.009)	1.42 (p=3.6e-06)	1.19 (p=0.0066)	1.30 (p=0.002)
7250653	CYP51	1.24	2.14	1.50	1.89

		(p=0.0116)	(p=1.8e-09)	(p=4.2e-05)	(p=6.7e-08)
7280610	UDP-Galactose-4-epimerase	1.65 (p=0.0006)	1.50 (p=0.0042)	1.36 (p=0.0243)	1.53 (p=0.0023)
7291805	Dehydrodolichyl diphosphate synthase	1.37 (p=0.0003)	1.36 (p=0.0005)	1.27 (p=0.0041)	1.17 (p=0.0447)
7317317	Squalene epoxidase	1.83 (p=2.4e-05)	2.27 (p=5.6e-07)	1.72 (p=0.0001)	1.70 (p=0002)
7329323	Solute carrier 39 (zinc transporter) member 4	2.81 (p=2.5e-10)	1.89 (p=9.9e-08)	1.65 (p=2.1e-06)	1.68 (p=1.2e-06)

To be considered part of a grouping, genes must have had a physiologic state $p < 0.05$ and at least one tissue simple effect $p < 0.01$. Reported p-values are the tissue simple effect p-values. The tissue simple effect p-value represents the comparison between Lactation and Control in the corresponding tissue. For the purposes of assigning patterns, the significance cutoff for the remaining tissues' simple effect p-values was set to $p < 0.05$. Abbreviations used as in Table 3.1. *Gene is at the Extended confidence level.

Table 3.3. Top three pathways for overrepresentation in each tissue

Liver

Pathway	Fisher's Exact test p-value	BH p-value	Members
Biosynthesis of steroids	3.50E-07	5.80E-05	↑Dhcr7, ↑Fdft1,↑Hmgcr,↑Idi1, ↑Lss, ↑Mvd, ↑Sqle
LXR/RXR Activation	1.57E-04	1.30E-02	↓Abcg5, ↓Abcg8, ↑Acaca, ↑Cyp7a1, ↓Hadh, ↑Hmgcr, ↓Lcat
Pentose/Phosphate Pathway	3.01E-04	1.67E-02	↑Aldoc, ↑G6pd, ↑Gpi, ↓H6pd, ↓Pgm5

Duodenum

Pathway			
Biosynthesis of steroids	1.13E-10	1.74E-08	↑Dhcr7, ↑Fdft1, ↑Fntb, ↑Hmgcr, ↑Idi1, ↑Lss, ↑Mvd, ↑Sc5dl, ↑Sqle
Androgen and estrogen metabolism	1.35E-04	1.04E-02	↑Ftsj1, ↑Hsd11b1, ↑Hsd17b7, ↑Nsdhl, ↓Srd5a2, ↓Ugt2b7
Glycerolipid metabolism	1.46E-03	7.50E-02	↓Cel, ↓Clps, ↓Glb1l2, ↓Pnlip, ↓Pnliprp1, ↓Pnliprp2

Jejunum

Pathway			
T-Cell receptor signaling	1.68E-07	2.78E-05	↓Bmx, ↓Cd8b, ↓Grap2, ↓Itk, ↓Malt1, ↓Pik3cg, ↓Pik3r1, ↓Ppp3cc, ↓Prkcq, ↓Ptprc, ↓Rasgrp1, ↓Trb

Pathway			
CD28 signaling in T helper cells	1.17E-05	6.74E-04	↓Grap2, ↓Itk, ↓Malt1, ↑Mapk9, ↓Pik3cg, ↓Pik3r1, ↓Ppp3cc, ↓Prkcq, ↓Ptprc, ↓Trb
TR/RXR activation	1.23E-05	6.74E-04	↑Fga, ↑Gh1, ↓Klf9, ↑Ldlr, ↓Pik3cg, ↓Pik3r1, ↓Thra, ↓Thrb, ↓Thrsp

Ileum

Pathway			
Biosynthesis of steroids	9.42E-08	1.71E-05	↑Cyp24a1, ↑Dhcr7, ↑Fdft1, ↑Hmgcr, ↑Idi1, ↑Mvd, ↑Sc5dl, ↑Sqle
TR/RXR activation	3.16E-07	2.86E-05	↑Fga, ↑Gh1, ↓Klf9, ↑Me1, ↓Nfcor2, ↓Pck1, ↓Pik3c2b, ↓Pik3cg, ↓Pik3r5, ↓Ppargc1a, ↓Thra, ↓Thrb
Thrombopoietin signaling	4.85E-05	2.53E-03	↓Irs2, ↓Pik3c2b, ↓Pik3cg, ↓Pik3r5, ↓Plcg1, ↓Plcg2, ↓Prkce, ↓Prkcq

The pathways with the three lowest p-values from Fisher's exact test for pathways from IPA's "Canonical Pathways" database for each tissue are reported here along with the corresponding p-value from a right-tailed Fisher's exact test, the Benjamini-Hochberg (BH) adjusted p-value, and the genes within the pathway that showed differential expression within the respective tissue (Members). Arrows in the Members column indicate the direction of change in lactating animals relative to controls.

Table 3.4. ATP-Binding Cassette (ABC) transporters detected to be differentially expressed between Control and Lactating dams.

Gene Symbol	Transcript Cluster ID	RL	RD	RJ	RIL	p<0.01
	7065486	0.888	0.931	0.99	0.8	
Abca3		(p=0.138)	(p=0.329)	(p=0.874)	(p=0.007)	IL
	7083965	0.567	0.885	0.942	0.763	
Abca8a_predicted*		(p=7.0E-04)	(p=0.273)	(p=0.744)	(p=0.054)	L
	7250393	0.63	0.663	0.652	0.697	
Abcb1a		(p=0.001)	(p=0.002)	(p=0.002)	(p=0.005)	L,D,J,IL
	7089622	0.818	1.065	0.938	0.672	
Abcc5		(p=0.021)	(p=0.476)	(p=0.471)	(p=8E-05)	IL
	7051110	0.886	0.852	0.884	0.73	
Abcc6		(p=0.109)	(p=0.03)	(p=0.089)	(p=2.0E-04)	IL
	7254219	0.949	1.55	1.049	1.203	
Abcg2		(p=0.666)	(p=6.0E-05)	(p=0.613)	(p=0.049)	D
	7303106	0.321	0.762	0.901	0.656	
Abcg5		(p=3.0E-07)	(p=0.095)	(p=0.515)	(p=0.006)	L, IL
	7294657	0.297	0.67	0.803	0.592	
Abcg8		(p=4.0E-06)	(p=0.039)	(p=0.261)	(p=0.005)	L, IL

The p<0.01 column indicates in which tissues a change was detected. Abbreviations are as defined for Table 3.1. *Gene is at the Extended level of confidence.

Additional files

Additional File 3.1: Statistical analysis and statistical pattern matching results (Statistical_Analysis_and_Statistical_Pattern_Matching_Results.txt)

Results from statistical pattern matching are reported as a .txt file. Results are reported for all 14,129 genes considered for statistical analysis, but genes that were not differentially expressed in any tissue are assigned a pattern of zero. Transcript_ID is an identifier associated with the gene analyzed. The gene assignment entry was taken from the Affymetrix annotation file and displays which gene is associated with the given transcript ID. Tissue p, physiological state p, and tissue*physiological state p correspond to the p-values associated with the tissue main effect, physiological state main effect, and the interaction, respectively. p Liver, p duodenum, p jejunum, and p ileum represent the p-values associated with the corresponding simple effects Pattern indicates in which pattern a gene is detected. A zero indicates no differential expression. Otherwise, patterns are assigned as described in the Methods. Ratio liver, ratio duodenum, ratio jejunum, and ratio ileum represent the ratio of the mean of samples from lactating animals to the mean of samples from controls utilizing the untransformed microarray data.

Additional File 3.2: Benjamini-Hochberg false discovery rates (Benjamini_Hochberg_False_Discovery_Rates.txt)

False Discovery Rate corrections [102] for all genes studied is presented in .txt format. Transcript_ID is an Affymetrix Identifier for each gene. False discovery rates are provided for the tissue*physiological state interaction term and the overall physiological state effect. Q-values were calculated as the number of false positives expected by chance (uncorrected p* total number of tests) divided by the total number of results with an equal or lower p-value.

Additional File 3.3: Histograms of p-values (Histograms_of_p_values.jpg)

Histograms for A) tissue effect p-values, B) physiological state effect p-values, C) physiological state*tissue interaction p-values, and pairwise comparison p-values for D) the liver, E) duodenum, F) jejunum, and G) ileum presented as a .ppt file. While a large tissue effect was observed, a visible treatment effect (control vs. lactation) was also observed.

Additional File 3.4: Volcano plots (Volcano_plots.pdf)

Volcano plots comparing the $\log_2$ fold changes (reported as mean untransformed lactating intensity divided by untransformed mean control intensity) against the calculated pairwise comparison p-value for each individual tissue in .pdf format. Volcano plots are for A) Liver, B) Duodenum, C) Jejunum, and D) Ileum. Each tissue responded differently to lactation. The blue line indicates the significance cutoff of $p < 0.01$. The number of differentially expressed genes were 420 in the liver, 337 in the duodenum, 402 in the jejunum, and 523 in the ileum, when an overall treatment main effect p-value cutoff of $p < 0.05$ was incorporated. Of

82

particular note is a series of genes that were strongly downregulated in the

duodenum (Additional File 3.1 pattern -100; discussed in Results.)

Additional File 3.5: DAVID output (DAVID_output_file.txt)

DAVID was utilized to test specific patterns from statistical pattern matching.

Patterns were chosen based on similarity between tissues and were "Up in All

Tissues", "Down in all tissues", "Up in all parts of small intestine", "Down in all

parts of small intestine", "Up only in liver", "Down only in liver", and "Down only in

duodenum". Table truncated from output in DAVID. Genes listed by Affymetrix

transcript cluster ID, which may be referenced in Additional File 3.1.

Additional File 3.6. Genes with decreased mRNA in all tissues

(Genes_with_decreased_mrna_all_tissues.pdf)

To be considered part of a grouping, genes must have a physiologic state p<

0.05 and at least one tissue simple effect p <0.01. Reported p-values are tissue

simple effect p-values and represent the comparison between Lactation and

Control in the corresponding tissue. For the purposes of assigning patterns, the

significance cutoff for the remaining tissue simple effect was set to p <0.05.

Abbreviations used as in Table 3.1. *Gene is at the Extended confidence level.

Additional File 3.7: Biosynthesis of sterols in liver

(Biosynthesis_of_sterols_in_liver.jpg)

Image from IPA representing the "Biosynthesis of Sterols" in the liver as a .jpg file. Numbering system for enzymes in the pathway is taken from KEGG [160]. Components of the cholesterol biosynthetic pathway include 1.1.1.34 (Hmgcr), 2.7.1.36 (Mvk), 2.7.4.2 (Pmvk), 4.1.1.83 (Mvd), 5.3.3.2 (Idi1), 2.5.1.21 (Fdft1), and 1.14.99.7 (Sqle), 5.4.99.7 (Lss), and 1.3.1.21 (Dhcr7). Red shading indicates increased mRNA during lactation from the corresponding gene.

Additional File 3.8: Biosynthesis of sterols in duodenum (Biosynthesis_of_sterols_in_duodenum.jpg)

Image from IPA representing the "Biosynthesis of Sterols" in the duodenum as a .jpg file. Numbering system for enzymes in the pathway is taken from KEGG [160]. Components of the cholesterol biosynthetic pathway include 1.1.1.34 (Hmgcr), 2.7.1.36 (Mvk), 2.7.4.2 (Pmvk), 4.1.1.83 (Mvd), 5.3.3.2 (Idi1), 2.5.1.21 (Fdft1), and 1.14.99.7 (Sqle), 5.4.99.7 (Lss), and 1.3.1.21 (Dhcr7). Red shading indicates increased mRNA during lactation from the corresponding gene.

Additional File 3.9: Biosynthesis of sterols in jejunum (Biosynthesis_of_sterols_in_jejunum.jpg)

Image from IPA representing the "Biosynthesis of Sterols" in the jejunum as a .jpg file. Numbering system for enzymes in the pathway is taken from KEGG [160]. Components of the cholesterol biosynthetic pathway include 1.1.1.34 (Hmgcr), 2.7.1.36 (Mvk), 2.7.4.2 (Pmvk), 4.1.1.83 (Mvd), 5.3.3.2 (Idi1), 2.5.1.21

84

(Fdft1), and 1.14.99.7 (Sqle), 5.4.99.7 (Lss), and 1.3.1.21 (Dhcr7). Red shading indicates increased mRNA during lactation from the corresponding gene.

Additional File 3.10: Biosynthesis of sterols in ileum (Biosynthesis_of_sterols_in_ileum.jpg)

Image from IPA representing the "Biosynthesis of Sterols" in the ileum as a .jpg file. Numbering system for enzymes in the pathway is taken from KEGG [160]. Components of the cholesterol biosynthetic pathway include 1.1.1.34 (Hmgcr), 2.7.1.36 (Mvk), 2.7.4.2 (Pmvk), 4.1.1.83 (Mvd), 5.3.3.2 (Idi1), 2.5.1.21 (Fdft1), and 1.14.99.7 (Sqle), 5.4.99.7 (Lss), and 1.3.1.21 (Dhcr7). Red shading indicates increased mRNA during lactation from the corresponding gene.

Additonal File 3.11. Genes regulated by Srebp proteins (Genes_Regulated_by_Srebp_proteins.pdf)

Genes that increase expression in Srebp-1a overexpressing mice and Srebp-2 overexpressing mice, and decrease expression in Scap knockout mice [87]. Overrepresentation analysis showed that genes in this list occurred more frequently than expected by chance in the lists of differentially expressed genes ($p < 1 \times 10^{-4}$ in each tissue.) Abbreviations used as in Table 3.1.   *Gene is at the Extended confidence level.

Additional File 3.12. Members of the Slc superfamily (Slcs.pdf)

85

Table displaying members of the Slc superfamily. The p<0.01 column indicates in which tissues a change was detected. Abbreviations are as defined for Table 3.1. * Gene is at the Extended level of confidence. [a]Substrates taken from the SLC tables database (http://www.bioparadigms.org/)

Additional File 3.13: Canonical pathways in the liver

(Liver_canonical_pathways.txt)

This file contains the canonical pathways in IPA and the corresponding $-\log_{10}$ p-values and the individual molecules that were detected as being in the list of overrepresented genes and part of each pathway (listed in the "molecules" column.) FDRs for the individual pathways are shown in a separate series of rows underneath the $-\log_{10}$ p-values. This table shows results utilizing the list of differentially expressed genes in the liver.

Additional File 3.14: Canonical pathways in the duodenum

(Duodenum_canonical_pathways.txt)

This file contains the canonical pathways in IPA and the corresponding $-\log_{10}$ p-values and the individual molecules that were detected as being in the list of overrepresented genes and part of each pathway (listed in the "molecules" column.) FDRs for the individual pathways are shown in a separate series of rows underneath the $-\log_{10}$ p-values. This table shows results utilizing the list of differentially expressed genes in the duodenum.

Additional File 3.15: Canonical pathways in the jejunum

(Jejunum_canonical_pathways.txt)

This file contains the canonical pathways in IPA and the corresponding $-\log_{10}$ p-values and the individual molecules that were detected as being in the list of overrepresented genes and part of each pathway (listed in the "molecules" column.) FDRs for the individual pathways are shown in a separate series of rows underneath the $-\log_{10}$ p-values. This table shows results utilizing the list of differentially expressed genes in the jejunum.

Additional File 3.16: Canonical pathways in the ileum

(Ileum_canonical_pathways.txt)

This file contains the canonical pathways in IPA and the corresponding $-\log_{10}$ p-values and the individual molecules that were detected as being in the list of overrepresented genes and part of each pathway (listed in the "molecules" column.) FDRs for the individual pathways are shown in a separate series of rows underneath the $-\log_{10}$ p-values. This table shows results utilizing the list of differentially expressed genes in the ileum.

Additional File 3.17: Thyroid pathway in liver (Liver_thyroid_pathway.jpg)

Images from IPA for the TR/RXR pathway for the liver. Red shading indicates increased mRNA amounts of the respective gene during lactation, and green shading indicates decreased amounts of mRNA.

Additional File 3.18: Thyroid pathway in duodenum (Duodenum_thyroid_pathway.jpg)

Images from IPA for the TR/RXR pathway for the duodenum. Red shading indicates increased mRNA amounts of the respective gene during lactation, and green shading indicates decreased amounts of mRNA.

Additional File 3.19: Thyroid pathway in jejunum (Jejunum_thyroid_pathway.jpg)

Images from IPA for the TR/RXR pathway for the jejunum. Red shading indicates increased mRNA amounts of the respective gene during lactation, and green shading indicates decreased amounts of mRNA.

Additional File 3.20: Thyroid pathway in ileum (Ileum_thyroid_pathway.jpg)

Images from IPA for the TR/RXR pathway for the ileum. Red shading indicates increased mRNA amounts of the respective gene during lactation, and green shading indicates decreased amounts of mRNA.

Additional File 3.21: RT-PCR primers (RT_PCR_primers.pdf)

Primer sequences for all genes analyzed by RT-PCR.

Chapter 4

Detection of Differential Alternative Splicing in the Lactating Rat

Introduction

Lactating dams have increased energy demands. Several physiologic changes occur in the animal to accommodate the need to provide for both themselves and their pups. These changes include increased cholesterol synthesis [9], increased zinc absorption [74], increased bile acid pool size [10], and increased hydrophobicity of the bile acid pool [10]. In addition, a number of hormone signaling pathways are altered, including those for insulin, thyroid, and leptin [19]. We have previously characterized the changes in mRNA expression in the liver and gastrointestinal tract (duodenum, jejunum, ileum) of the lactating dam compared to age matched virgin controls using an Affymetrix Rat Exon 1.0 ST microarray (Affymetrix, Santa Clara, CA) [164].

In addition to changes in gene expression, alternative splicing is also a mechanism for controlling gene function. To our knowledge, no work has been done to characterize changes in mRNA splicing during lactation. One tool that has been developed to characterize differences in splicing between two different groups is the Exon Array. In contrast to traditional 3'-expression arrays, probes in the Exon Array are designed to detect individual exons. This allows for measurements of individual exons, and a comparison of exon expression

between two groups relative to the corresponding gene expression allows for identification of genes that are differentially spliced between the groups.

In spite of the minimal work done to characterize splicing within the context of the lactating animal, the serum levels of several hormones that affect splicing are known to be altered. Lactating rats are hypoinsulinemic [19], hypoleptinemic [19], and hyperprolactinemic when compared to nonlactating animals. These pathways influence PI3K signaling [37, 38, 52, 54, 57, 58, 60, 61]. Insulin signaling can trigger differential splicing of protein kinase C β II through the PI3K pathway and ultimately through differential phosphorylation of SRp40, a member of the serine-arginine (SR) family of proteins which are often involved in the regulation of alternative splicing [68]. Likewise prolactin receptor activation leads to differential splicing of neuronal nitric oxide synthase (nNOS) in the rat anterior pituitary cell line, GH_3 [67]. Consequently, differential alternative splicing through differential phosphorylation of SR proteins downstream of PI3k signaling seems likely in lactating animals.

Several methods have been developed to analyze data from exon array chips. Two of these methods are ANOSVA [107] and MiDAS [97]. ANOSVA is a method based on a two-way ANOVA, utilizing the factor (i.e. source of variance) of interest and an exon effect as the two main effects. A significant interaction term between the effect of interest and the exon is considered a positive test. This method operates under the model $y_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+error$, where y_{ijk} is the measured expression level for probe k in probeset i in experiment j, μ is a

baseline measurement, α_i is the linear contribution of probeset i, β_j is the linear contribution of the other factor, and γ_{ij} is the interaction effect. In the presence of differential alternative splicing, the addition of a splicing event creates a situation where the sum of μ, αi, and βj do not fully explain the model, and the difference is explained in the interaction term. However, problems have been reported with this model, as an Affymetrix white paper discussing the method described that it "did not yield good performance" but specific flaws were not discussed [165]. Also, because the test is performed for the entire gene, the method does not give a direct measure of where the potential splice sites occur.

Another method used to detect differential alternative splicing is called MiDAS [97]. This method utilizes "gene normalized intensities" to identify where splicing has occurred. The gene normalized intensities are calculated as the ratio of an individual exon's expression to an estimate of the gene's overall expression. If a significant difference occurs between "gene normalized intensities" for a given exon, then that exon may be differentially spliced. Analysis is often done by utilizing a t-test on the gene normalized intensities [97]. In some cases, this method has reported low validation rates, as there are several well-known sources of false positives in exon microarrays. These include failure of a probe to hybridize to its target cDNA, hybridization to nontarget cDNAs, and a gene not being expressed in one of the treatment groups [97]. Because of these problems, adequate prefiltering before analyzing the data is a necessity [97].

We have utilized a combination of ANOSVA and MiDAS to detect differential splicing. We have characterized differential alternative splicing in the liver, duodenum, jejunum, and ileum of lactating rats compared to age matched virgin controls by running ANOSVA to identify differentially spliced genes, followed by MiDAS to identify differentially spliced exons.

Results

ANOSVA Overestimates p-values

The physiologic state*probeset interaction p-values were plotted as histograms

for each tissue (Figure 4-2). In the absence of any significant interactions (and

under the assumptions of the algorithm, in the absence of alternative splicing),

the interaction p-values would be expected to be normally distributed. However,

histograms of the p-values show that p-values were biased toward one (1) in

each tissue, although in the ileum, there was both a bias toward zero (0) and one

(1). This indicates that the ANOSVA method is in general, overestimating the p-

values, as the expected uniform distribution of nonsignificant results was not

present. Histograms of MiDAS p-values (Figure 4.3) appeared to follow a similar

trend, but the effect was much less pronounced, since there were multiple exons

for each gene.

Number of genes that tested positive in each tissue for alternative splicing

Because low validation rates have been reported for microarray experiments,

and validation can be a time-consuming and costly endeavor, we chose a

relatively stringent cutoff at the level of ANOSVA (FDR=0.05). In three of the four

tissues, very few genes (i.e., less than 20) were detected as positives (Tables

4.1-4.3). Given that an FDR of 0.05 indicates that approximately 1 in 20 genes

are false positives, and fewer than 20 genes appear in these lists, it is plausible

that some of these genes tested positive by chance. However, the ileum

93

displayed a much higher incidence of positive tests (Additional File 4.1). This is indicative of a unique event occurring within the ileum, as all tissues were analyzed using the same model.

UGTs Test Positive for Alternative Splicing

In both the liver and the ileum, the gene associated with the UDP glucuronosyltransferase 1 family polypeptide a (UGT1a) tested positive for alternative splicing. The UGT1a differ by usage of an alternative first exon. The splicing events that tested positive by the microarray were associated with probesets that match the alternative first exons. In the liver, probesets 6347739 and 6443473 tested positive for differential alternative splicing, while in the ileum, no individual probesets tested positive in MiDAS, although the ANOSVA for the gene gave a significantly low p-value. By using University of California, Santa Cruz (UCSC) genome browser [166, 167] to align the target sequence against the genome, we identified probeset 6347739 as being associated with the first exon of Ugt1a7c, and probeset 6443473 as being associated with the first exon of Ugt1a6.

Alternative Splicing of Abcg8

Initial analyses of the extended dataset prior to filtering out poorly expressed probesets resulted in a larger number of positive tests. Within this dataset, Abcg8 was flagged as a likely candidate for alternative splicing in the liver only (positive at FDR<0.05; Figure 4.4). RT-PCR detected a splice variant that was

present in lactating samples, and this variant was sequenced. (Additional File

4.2). Only one transcript variant was sequenced, although others were identified

in the RT-PCR products.

Discussion

A Series of Events Occur in the Gastrointestinal Tract That is Unique to the Ileum

The same algorithm was utilized to detect the possibility that lactation triggers changes in alternative splicing in each tissue. In three of the tissues, very few positive tests were detected. However, the ileum showed substantially more positive alternative splicing events (89 in the ileum, less than 20 in each of the other tissues). Although events such as alternative start site and stop site usage would also test positive as alternative splicing events, this indicated that a unique physiologic event was occurring in the ileum of lactating rats that was not occurring in the liver, duodenum, or jejunum. Based on differences in serum hormone levels reported in lactation, we had reason to hypothesize that levels of alternative splicing in the gastrointestinal tract were altered during lactation. Thus, changes in signaling via the PI3k/Akt pathway could affect splicing mechanisms. However, in three of the four tissues, the number of splicing events detected by the microarray chip was small. Alternative splicing events in any of these tissues in the lactating animal have not been characterized previously. Further work needs to be done to characterize what changes in the local environment of the ileum might trigger differences in alternative splicing that are unique to the ileum when compared to the other more proximal parts of the small intestine and the liver.

96

Abcg8 is Differentially Spliced in the Liver of Lactating Dams, but not in the Small Intestine

Our previous analyses [5] showed that Abcg8 is differentially expressed in the livers of lactating rats. mRNA and protein levels are significantly decreased, and functional heterodimer was not detectable in livers of lactating dams. Initial analyses of the microarray data for the detection of alternative splicing prior to filtering out poorly expressed genes indicated that Abcg8 might be differentially spliced in the liver of lactating animals. From the plots of probeset intensity, the probesets in the liver of lactating rats show a clearly different pattern of expression than in any other tissue or physiologic state group (Figure 4.4). We investigated the cause of the unique expression pattern in these livers. RT-PCR identified a transcript variant within the lactating liver samples. This variant was sequenced and found to be missing part of exon 4, all of exons 5,- 8, and part of exon 9. Blasting the sequence of the clone against the rat genome identified the best match was Abcg8, with several regions having 100% identity and others varying only by a single nucleotide mismatch or a gap. Comparing the determined sequence against NCBI's NM_130414 mRNA entry and NP_569098 protein entry, it was determined that this transcript most likely would not have been able to produce functional protein. In addition to the possibility that the transcript's product may not have been able to form a dimer with Abcg5, this transcript would likely not include many key features of the transporter, including

the Walker B motif in the nucleotide binding domain where ATP is bound. This result indicated that within the lactating liver, incomplete transcripts of Abcg8 were being formed, and these transcripts may not be able to form the functional protein, providing further description for the lack of functional Abcg8 protein detected in the liver of lactating dam. This, in addition to the substantially decreased amounts of Abcg8 mRNA, suggested that control of Abcg8 levels in lactating dams happened at least in part at the level of transcription. As we have described in Chapter 3, loss of Abcg5/g8 requires that cholesterol be secreted into bile through an alternative mechanism rather than through transport mediated by the Abcg5/Abcg8 heterodimer. Abcg5 was not detected to be differentially spliced in the same manner, but failure to produce functional Abcg8 interferes with the formation of functional heterodimer [168]. Abcg8 was removed from the final analysis in liver due to its already low expression in liver samples from lactating animals.

The Gene Associated with Members of the Ugt1a Family of Enzymes Shows Differential First Exon Usage in the Microarray Data in Both the Liver and the Ileum

mRNA associated with the Ugt1a proteins tested positive for differential alternative splicing in the liver and ileum. The Ugt1a locus codes for a family of proteins that differ in their amino termini, but have consistent carboxy termini. This is accomplished through utilization of a variable first exon, followed by the

98

consistent usage of exons 2-5. These enzymes are responsible for transferring glucuronic acid to the substrate, which generally creates a water soluble, biologically inactive product. These enzymes are important in both elimination of endogenous substrates and various drugs. A review of many example drug substrates can be found in Kiang et al [169].

Because the microarray makes measurements on each individual exon, differences in the usage of the differential first exon can be identified when testing for alternative splicing. In both the liver and the ileum, the Ugt1a locus was identified as a site of "differential alternative splicing", and further analyses revealed that the detected sites were located in the alternative first exons. In the liver, Ugt1a6 and Ugt1a7 were identified as having higher expression of the corresponding first exon present in the lactating samples compared to controls, while in the ileum, no forms were found differentially expressed by MiDAS, but Ugt1a3 and Ugt1a5 appeared to have slightly decreased expression of mRNA from the respective first exons in lactating dams (Figure 4.5). However, the constituitive exons 2-5 did not display a change in expression. These results suggested that drug metabolism may be altered in lactating dams. For example, Ugt1a6 can metabolize acetaminophen. The results from Ugt1a6 were consistent with reports of transgenic mice expressing the human UGT1A locus having higher levels of UGT1A6 post-partum [170]. Also, Ugt1a6 mRNA and protein have been shown to be increased in lactating rats and in rats treated with ovine prolactin [94], indicating that our results regarding Ugt isoforms were

99

consistent with previous literature.   Further studies need to be performed to elucidate how differences in Ugt levels actually influence drug metabolism.

Other Genes Detected to be Differentially Spliced and Biological Roles

Because the ileum showed a substantially higher number of splicing events compared to the other three tissues, we chose to focus on it for further analyses. Additional File 4.1 displays genes that tested positive in the ileum and corresponding physiologic state*probeset_ID interaction p-values as calculated by Partek. *Angiotensin converting enzyme 2 (Ace2)* showed the lowest p-value for alternative splicing in ileum using the two-way ANOVA model in Partek (Figure 4.6). MiDAS results, in conjunction with plotting exon intensities as a function of chromosomal position in both groups, revealed that probeset 5728701, which maps to the 3' end of exon 9 in the UCSC genome browser, was a possible site of differential alternative splicing, as a drop in both groups occurred, but the magnitude of the drop was greater in the control group. This would indicate that proportionately fewer transcripts contained the exon in control than lactating animals. In the duodenum and jejunum, a large decrease in expression of the probeset was observed, but there was no statistically significant difference in the gene normalized intensities between the groups. These results indicate that the splicing event was unique to the ileum.

Ace 2 catalyzes conversion of angeniotensin II into the inactive heptapeptide Ang1-7. Active angiotensin II is responsible for triggering increased blood pressure [171] and secretion of aldosterone, which leads to sodium retention [172]. The degradation of angiotensin II prevents further activity, meaning that ACE2 plays a key role in regulating blood pressure and sodium levels. Furthermore, the heptapeptide Ang1-7 can inhibit angiotensin II activity [173].

Using BLAT [174] against the Nov. 2004 (Baylor 3.4/rn4) assembly [175] to map the location of probeset 5728701's target sequence in the UCSC genome browser did not reveal any ESTs for Ace 2 that were spliced in the corresponding region. The most likely explanations for the results observed are that the probe hybridizes poorly, as marked by the large decrease in intensity relative to other probes in the same gene in all parts of the small intestine, or that the splicing event was real and was unique only to the ileum. The region corresponding to probeset 5728701 matches the mRNA refseq entry NM_001012006 from nucleotides 1301 to 1326. This was consistent with amino acids 434-442 in the translated protein, which can also be viewed in the NCBI refseq entry NM_001012006 or in the protein entry NP_001012006. This region was included with the peptidase region from amino acids 22-612, but was not associated with any active sites within the protein entry, which map to amino acids 345-348,374-375,378,382,401-402,449,503,505,512, and 515. Consequently, if this splice form interfered with function of the full protein, it would likely be from a conformational change through missing residues near the active sites; the

closest active site was threonine 449. Because this splice form retains its active sites and its transmembrane domain, the splice variant may still be biologically active, however this requires independent verification.

DAVID [105] was used to identify similarities amongst genes that were differentially alternatively spliced in the list of 89 genes detected in the ileum. Amongst these, four members of the renin angiotensin (p=5.4X10-4) pathway were detected, including the previously mentioned *Ace2*. Figure 4.7 shows the KEGG renin angiotensin pathway, as displayed from within DAVID. Members marked with stars were members of the renin angiotensin pathway. The other three members were alanyl (membrane) amino peptidase (*Anpep*), leucyl/cystinyl peptidase (*Lnpep*), and membrane metallo endopeptidase (*Mme*).

Seventeen members of the SP_PIR_Keywords group "alternative splicing" tested positive for differential alternative splicing (p=$4.1X10^{-3}$) (Table 4-4). Amongst these was the Ugt1a locus, which produces different proteins based on usage of an alternative first exon and is discussed in detail above. These are proteins that have been previously associated with alternative splicing.

In the liver, probesets associated with Ugt1a6 and Ugt1a7 were detected as differentially spliced. This suggested possible differences in improved ability to conjugate phenols, such as acetaminophen. In the rat, Ugt1a6 and Ugt1a7 are co-regulated through Ahr [95], suggesting that these results may have been a

response to the same mechanism. However, this result is not directly

transferable to humans, as humans do not express Ugt1a7 in the liver [176]. This

result was consistent with previous reports that Ugt1a6 mRNA and protein ares

increased during lactation and in response to prolactin [94].

Using a network drawn around Ahr in Ingenuity Pathways Analysis (Ingenuity

Systems, www.ingenuity.com), we found that four genes differentially expressed

in the liver would be influenced by Ahr. Two changes were consistent with

increased Ahr activity: increased Cyp1a1 [177] and decreased Serpina7 [178].

The other two changes that would be influenced by Ahr activity occured in the

direction opposite of the expected change. Hmgcr [179] transcription was

increased [164] and Phophoenolpyruvate carboxykinase 1 (Pck1, also known as

Pepck), a key regulator in gluconeogenesis, transcription was decreased [178].

The inconsistent changes were easily explained by the change in the lactating

rats' metabolic state, as discussed in Chapter 3.

RT-PCR Validation of Alternative Splicing Events

The following factors should be considered when determining which alternative

splicing events are likely true positives. First, does the exon in question show a

substantial difference in intensity and gene normalized intensity from the other

exons? Low variance across the measurement could lead to a statistically

significant result, but if the cause of that low variance was not a direct result of

the treatment, then the answer would be a false positive. Relative expression of the exon to the other exons in the gene may be an indication of actual alternative splicing, since it should be expressed at a lower level than constitutively expressed exons, although probe affinity should be accounted for when comparing across exons.

Another factor to take into account when choosing genes for which to validate splicing events is the presence of existing known splicing events [97]. The existence of known exons indicates that the transcript variant in question exists, and the only thing needed to be validated is the relative level of transcripts in each treatment. In some cases, the experimenter may be interested in novel splice variants or the animal model's transcriptome has been poorly sequenced. In these cases, the experimenter may need to consider if consensus or alternative acceptor and donor sites exist within the sequence. The donor and acceptor sites indicate the 5' and 3' ends of the splice site, respectively [180]. If the region of interest is on the 5' or 3' end of the message, alternative initiation and poly adenylation sites may need to also be considered.

Conclusions

The results of this study are consistent with previous findings that Ugt1a6 expression is increased in the liver of lactating dams. Our results indicated that Ugt1a7 may be regulated in a similar manner. We have detected an mRNA transcript variant of Abcg8 in the liver that further indicates that control of the levels of functional Abcg5/Abcg8 heterodimer are controlled at the level of mRNA. Our results also suggested possible differential alternative splicing of four members of the renin-angiotensin pathway, especially Ace2, which is key in inhibiting the renin-angiotensin pathway. Further work needs to be performed in order to validate exactly how splicing events are regulated in the gastrointestinal tract in lactating animals when compared to virgins. For example, the ileum was more responsive to changes in alternative splicing than the liver, duodenum, or jejunum. Current methods for identifying differential alternative splicing, particularly ANOSVA and MiDAS need to be refined to be applicable to the microarray dataset, as p-values did not form a uniform distribution for non-significant tests, indicating that assumptions within the model were violated. We have specified a minimal difference filter in order to attempt to rectify some of the false positives. Finally, RT-PCR using primers flanking probable splice sites for genes that were strong candidates for differential splicing, such as Ace2 in the ileum needs to be performed. Validating the RT-PCR results will describe how

well the additional filters work, as well as provide insights regarding the biology of

the lactating animal, as differential splice variants may play different biological

roles.

Materials and Methods

Animals

The data used here were obtained as described in Chapter 3 and published recently [164]. Briefly, sixteen Sprague-Dawley rats at day 10-11 of lactation and sixteen age matched virgin controls were obtained from Harlan (Indianapolis, IN). All animals were sacrificed at 16 h (10 h of light on a 12 hour light/dark cycle; 4 PM), and the liver, duodenum, jejunum, and ileum removed for total RNA extraction from each tissue. The integrity of all RNA samples was verified by an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Animal protocols were conducted in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee of the University of Kentucky. Each rat was assigned to one of four pools within the respective physiologic state (treatments; four control pools and four lactating pools.) Pooled RNA samples, consisting of the RNA from the four rats within the same group, were created for each tissue, with individual rats composing the pools consistent across tissues. resulting in the use of 32 chips (4 tissues X 2 "treatments" X 4 pools). Samples were prepared and processed according to the manufacturer's instructions by the University of Kentucky Microarray Core Facility (Lexington, KY)

Data Filtration

The data were filtered in order to reduce the risk of false positives. Affymetrix

Expression Console software (Affymetrix, Santa Clara, CA) was used to perform

the Robust Mutichip Averaging (RMA) [98, 99] algorithm on each tissue

separately on the "Core" dataset.   For example, the eight chips associated with

the liver were analyzed at the same time, but were analyzed separately from the

eight chips associated with the duodenum. During these calculations, Detection

Above Background (DABG) values were calculated. A DABG of less than 0.01

was considered a positive test, and probesets (treated as exons for the purposes

of the calculations) with DABG values of less than 0.01 were considered present

on the respective chip. If a probeset was present in at least three of four chips in

control or lactating groups, the exon was considered present in that group. If a

probeset was not considered present in both groups, it was removed from

analysis for determining differential splicing. A column determining whether a

given probeset was present for a given group (control or lactation) (in which case

a value of "1" was assigned) or absent (in which case a value of "zero" was

assigned) was added to the spreadsheet. A second presence call was

generated with a value of "1" indicating that the probeset was present in at least

one group, and a value of "0" indicating the probeset was present in neither

group. This was repeated for probeset intensity, where a cutoff of the log2 RMA

intensity was 4 rather than applying a DABG cutoff, and the presence/absence

assignment was done in the same way. The Affymetrix Rat Exon 1.0ST

annotation file (raex-1_0-st-v1-na30.rn4) was also used to identify genes that

have been characterized as risks of cross-hybridizing to nontarget cDNAs. Any

exon identified as a risk was removed from analysis. Affymetrix has identified probesets that may have issues with cross-hybridizing to cDNA other than the intended target. In the Affymetrix annotation file, a value of "1" in a column labeled "Crosshyb_type" is assigned to probesets that are not likely at risk to cross-hybridize to nontarget cDNA. Any value other than "1" in the annotation file under the column "Crosshyb_type" was determined to be a risk for cross-hybridization. In our dataset, a column containing a "1" for the row if the probeset in question was determined not likely to cross-hybridize and a "0" if the probeset was likely to cross-hybridize was created. The product of the three presence/absence calls was used to determine if a probeset was present or absent. Presence calls were made for each group and overall. A gene was considered present overall if both the DABG cutoff and the intensity cutoff indicated half of the probesets were present in both control and lactation.

JMP Genomics (SAS Institute Inc, Cary, NC) was used to summarize each of the probeset_IDs into transcript clusters (genes). The sum of present exons in control or lactating samples were calculated for each gene. If 50% of the exons in a given gene were not expressed in either control or lactation, the gene was removed from analysis. An intensity cutoff at the gene level was applied in a similar manner to the intensity cutoff applied at the exon level. If a gene expressed an intensity of at least four in three of the four chips within a group, it was considered present for the group; otherwise, it was considered absent for that group. This time, a gene needed to be present in both groups to be

109

considered present overall. These steps were repeated for each tissue.

Presence/absence calls were used to generate a text file that contained a list of

probesets to be used in the final analysis.

Detection of Differentially Spliced Genes

Affymetrix .CEL files were uploaded into Partek Genomics Suite ©2011 (Partek

Incorporated, St. Louis, MO) [181], and the RMA algorithm was utilized to treat

the data. After implementing the RMA algorithm, a list of probesets to be utilized

according to the filtration procedures described above was used to filter the

dataset. Each tissue was loaded separately into Partek, and the RMA algorithm

was performed separately for each tissue. Partek was then used to run a two

way ANOVA treating physiologic state and exon as the two main effects.   The

physiologic state*exon interaction term was used as an estimate for the likelihood

that a gene was differentially spliced within lactating animals compared to virgin

controls. A false discovery rate (FDR) was calculated by Partek and a value of

0.05 was used as a cutoff for a positive alternative splicing test in each tissue.

Detection of Differential Splice Sites

RMA values of exon and gene intensities from the RMA algorithm performed in

the Affymetrix Expression Console software was used to calculate "gene

normalized intensities". The gene normalized intensity was defined as the RMA

calculated intensity value of a probeset_ID divided by the intensity value of the

corresponding gene level intensity. A t-test for these values with respect to

physiologic state was performed for each exon within a gene that tested positive, and this procedure was repeated for each tissue. A p < 0.01 in MiDAS was considered a positive test for differential splicing of a given exon. The t-tests were performed in Microsoft Excel (Microsoft Corporation). Graphs of exon level intensities as a function of treatment and exon for genes that tested positive were also investigated in Partek. Because low variance appeared to be an issue in some positive tests, we also specified that a minimal difference of 0.5 on a $\log_2$ scale between mean exon intensities be present. The minimal difference filter would filter out situations where poor probe expression occurred but failed the presence/ absence filter. There were other cases in which the difference between exon expression appeared minimal when the data were displayed graphically (Figure 4-1), and consequently only positive tests that also displayed a large difference (0.5 on a $\log_2$ scale) were considered positive for the purposes of detecting the locations of potential splice sites.

RT-PCR and Sequencing of an Abcg8 Splice Variant

RT-PCR was performed on a Roche Light Cycler 280 instrument (Roche Applied Sciences, Manheim, Germany). The forward primer for Abcg8 was ATGGCTGAGAAGACCAAAGAGG, and the reverse primer was ATCATTGTCCTCAGTCGGGCTC. Samples were taken from the original RNA samples from the liver used for the microarray. RT-PCR products were separated on an agarose gel. A single RT-PCR product was extracted that was detected in the lactating sample, and this product was sequenced at MWG

biotech (Huntsville, AL). This portion of the study was carried out by Dr.
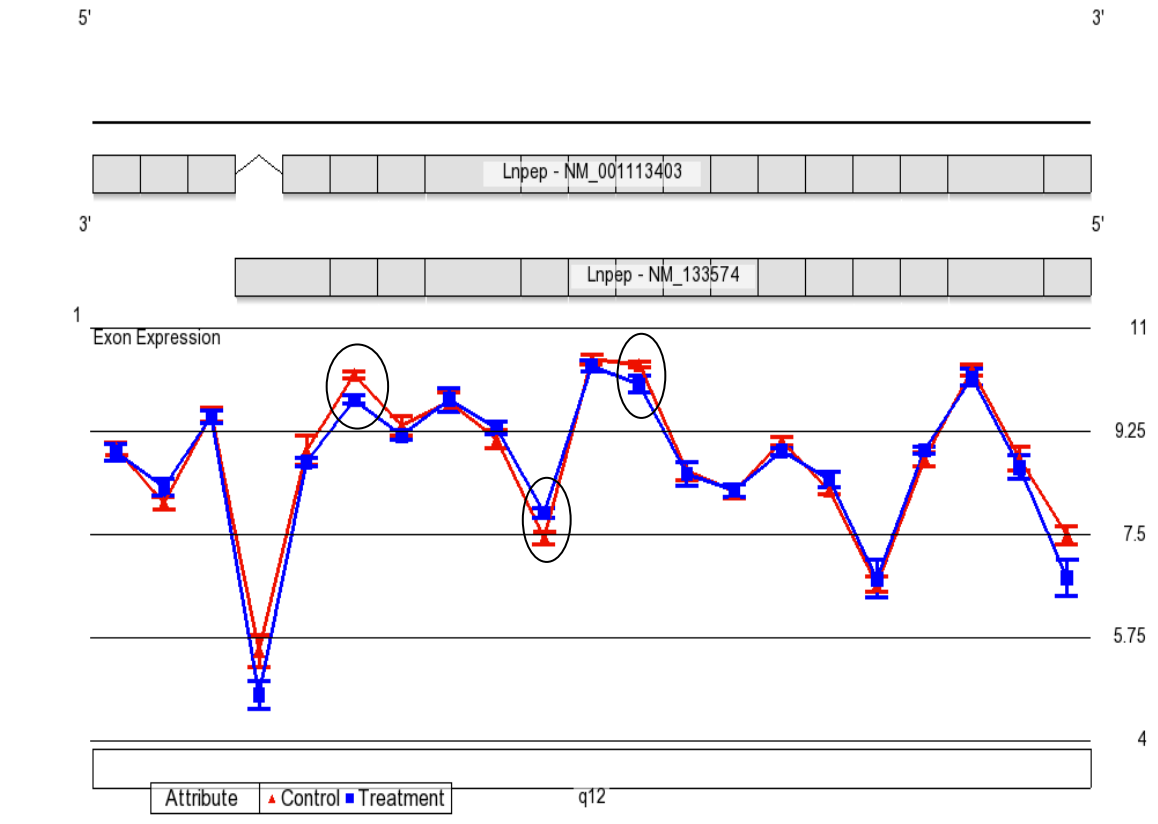
Tianyong Zhao in our laboratory.

Figure 4.1: A sample case of low variance leading to a likely false positive in MiDAS. Lnpep tests positive by ANOSVA, and three probesets test positive at p<0.01 in MiDAS. Although these probesets have low p-values due to low within-group variance, validation of an actual splicing event and its biological relevance at these sites would be difficult. The figure was extracted from Partek genomics suite, with probesets that test positive for MiDAS being circled. The circled exons were removed from the list of positive exons if a minimal difference of 0.05

between probeset intensities was enforced. Intensities (right axis) are plotted as

a function of chromosome position (bottom axis). Chromosome number is listed

on the left axis. The two Refseq transcripts are listed across the top

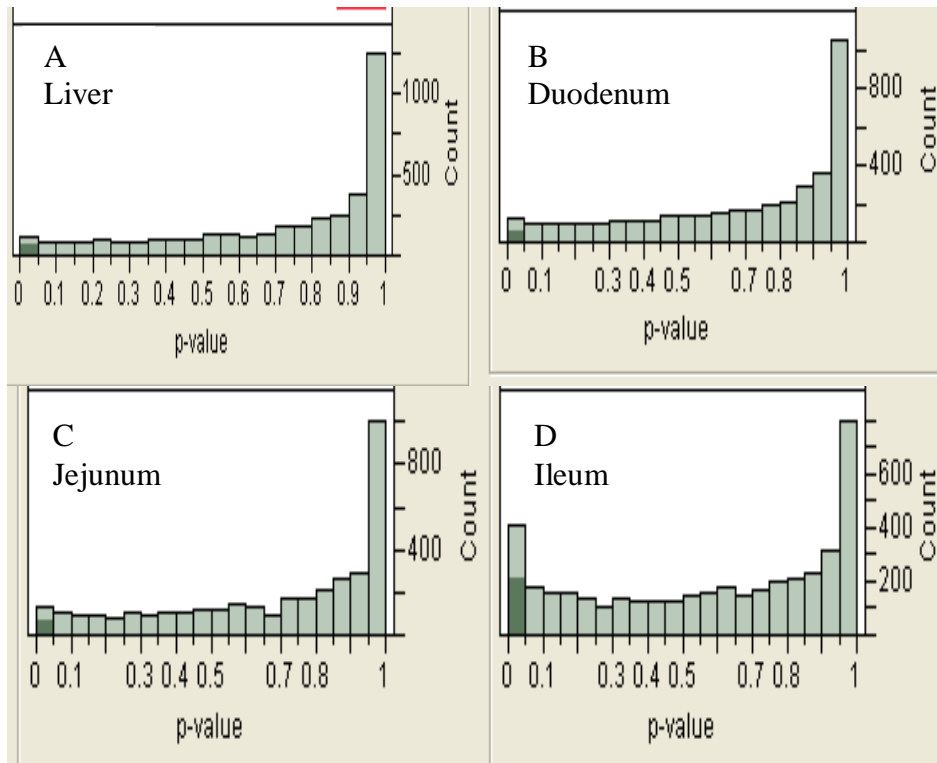(NM_001113403 and NM_133574), with boxes representing individual exons.

Figure 4.2: Histograms of alternative splicing p-values calculated by Partek. Values show a bias toward p=1.0, suggesting that underlying assumptions within the model are violated. On the x-axis are p-values and the y-axis represents the number of genes at the corresponding p-value. A. Liver, B. Duodenum, C. Jejunum, and D. Ileum. The darker shaded region represents $p<0.01$.
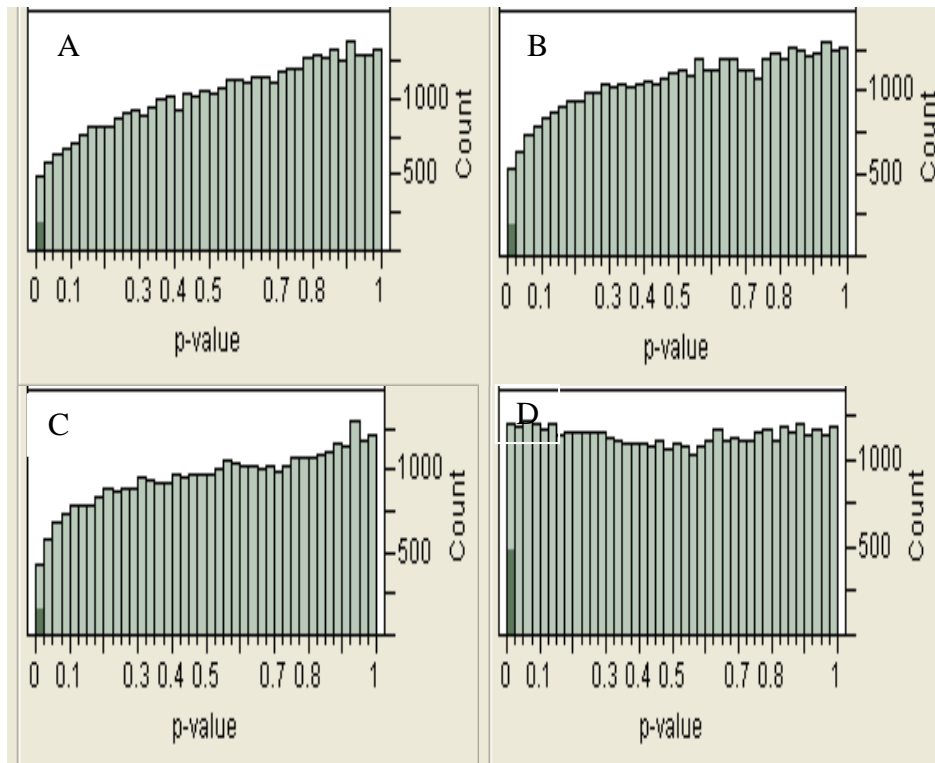
Figure 4.3: Histograms of MiDAS p-values for probesets detected as present on the microarray chips in the respective tissue (A. Liver, B. Duodenum, C. Jejunum, and D. Ileum). On the x-axis are p-values and the y-axis represents the counts of probesets at the corresponding p-value. In the liver, duodenum, and jejunum, an upward shift appears toward p=1, indicating that nonsignificant results were not uniformly distributed, suggesting some underlying assumptions regarding the MiDAS algorithm have been violated. Darker shaded regions represent p<0.01.

A
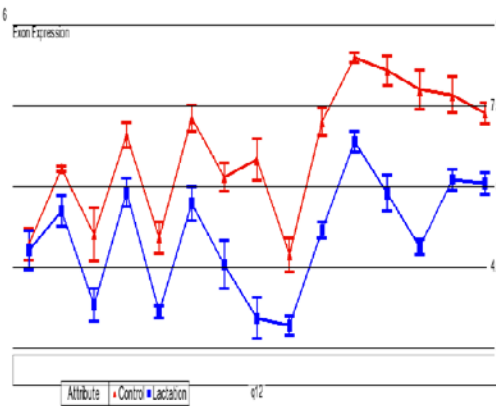
Liver

B

Duodenum

C

Jejunum

D

Ileum

117
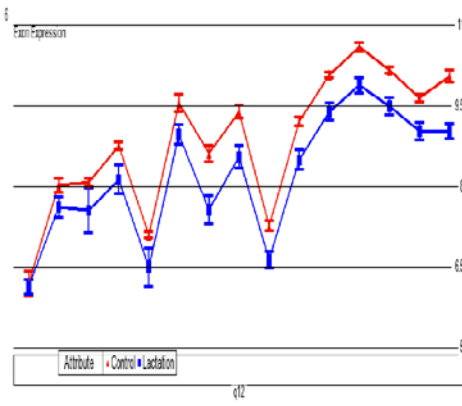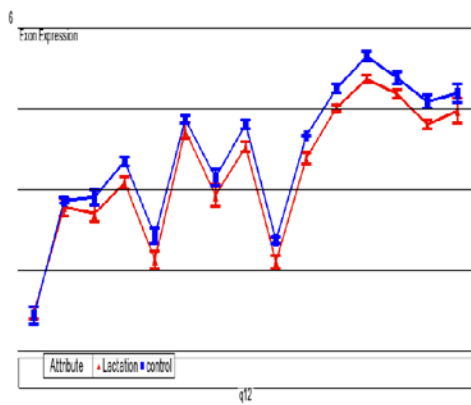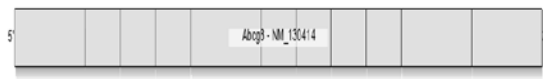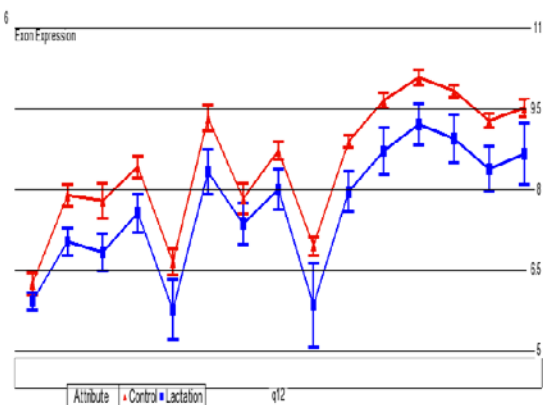
Figure 4.4: Abcg8 probeset expressionin A. Liver, B. Duodenum, C. Jejunum, and D. Ileum. Data compare probeset expression in lactation (blue line) against control (red line) before setting filters to remove absent exons and genes. The liver responds substantially differently to lactation at the level of individual probesets than the other tissues. This has been validating using RT-PCR. . Intensities (right axis) are plotted as a function of chromosome position (bottom axis). The left axis lists the chromosome number. Along the top is the NCBI refseq entry for Abcg8, with boxes representing individual exons.
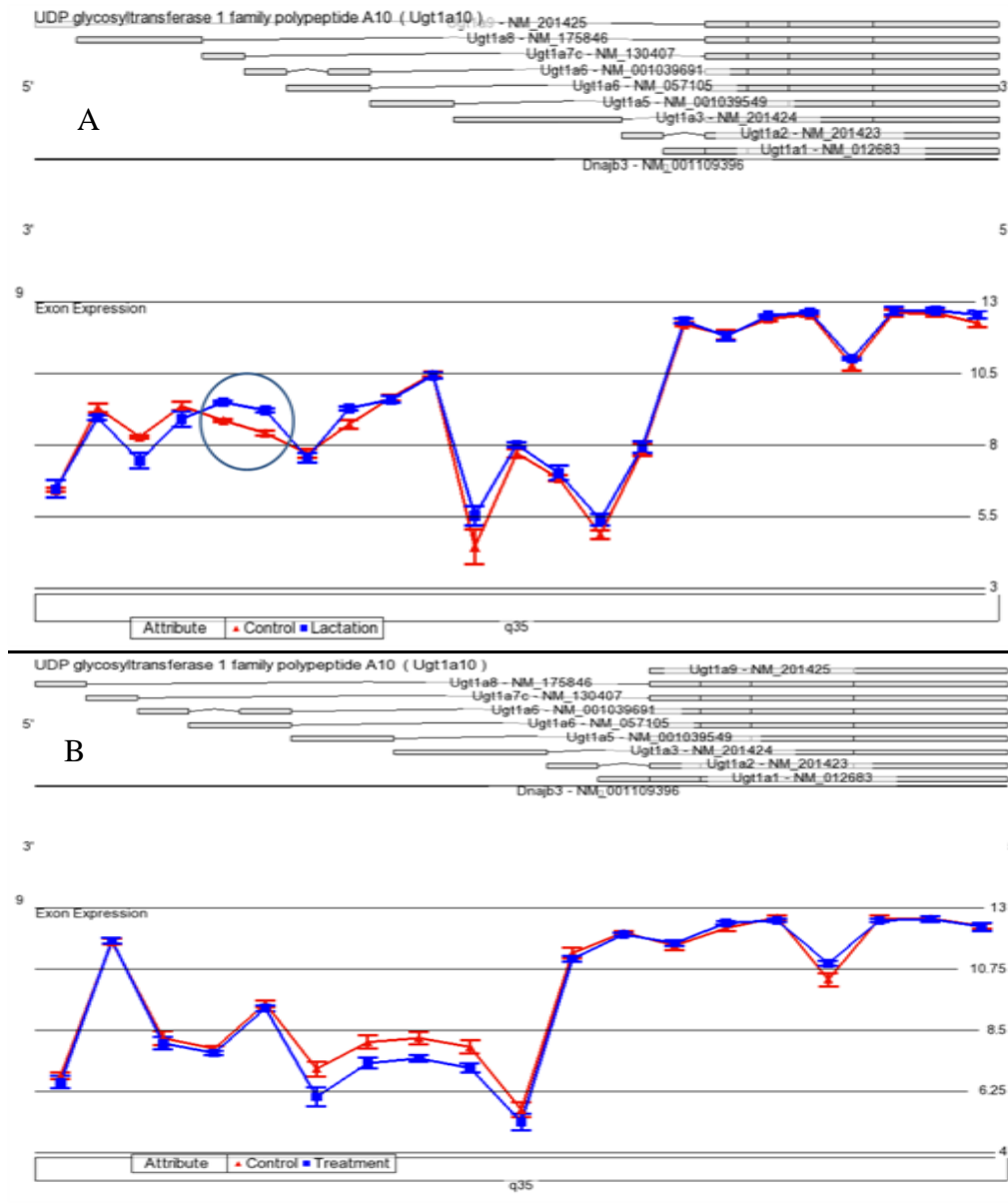
Figure 4.5: Probeset expression of the Ugt1a locus in liver (A) and ileum (B). In the liver, genes that tested positive by MiDAS at p<0.01 are circled. In the ileum, no probesets passed this test. In the liver, exons associated with Ugt1a6 and

Ugt1a7 show increased expression during lactation (blue line) compared to virgin

controls (red line).
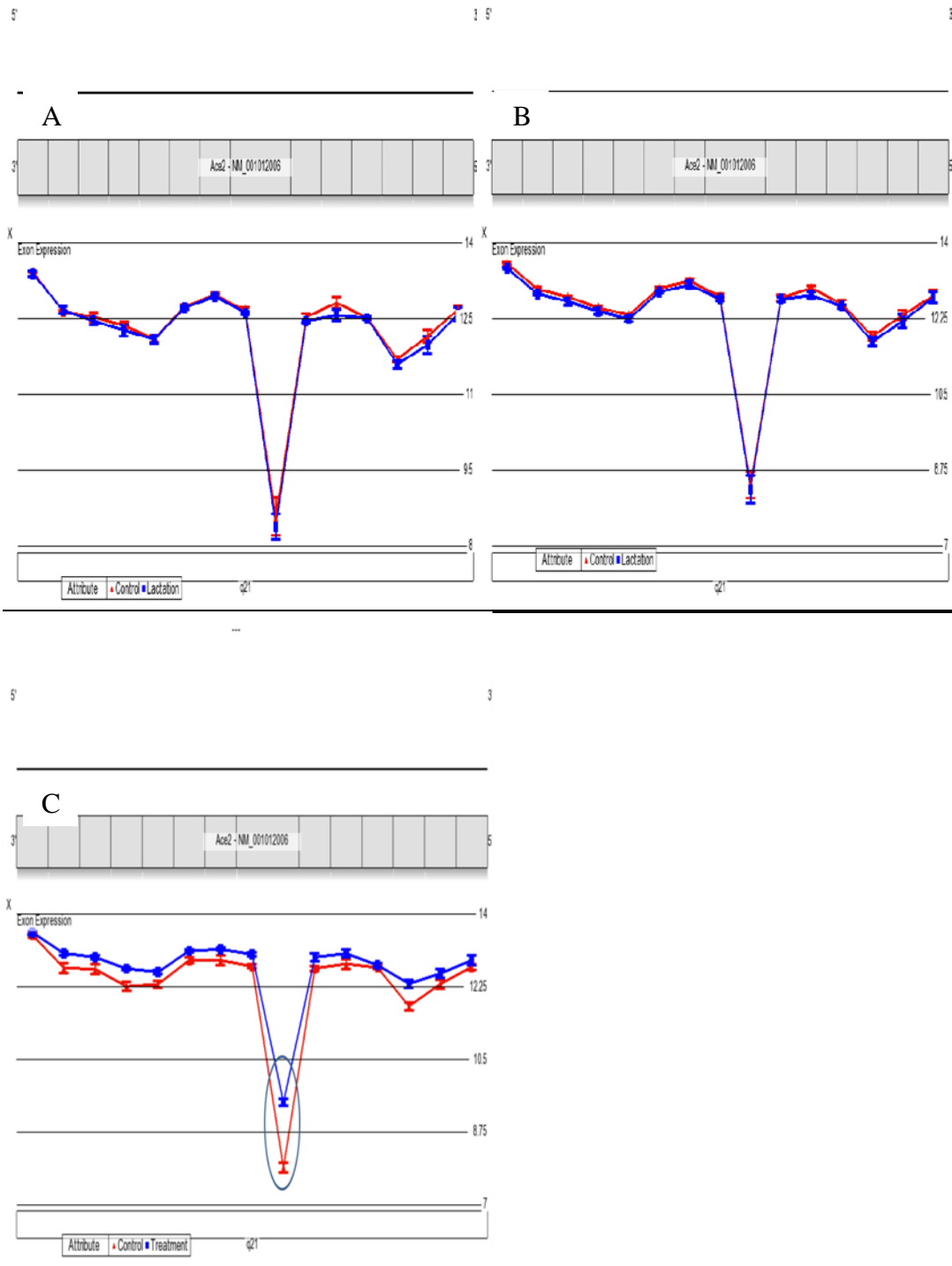
Figure 4.6: Ace2 probeset expression in duodenum (A), jejunum (B), and ileum

(C). Individual probeset intensities are plotted as a function of chromosome

position as indicated by Partek. Each tissue displays a substantially lower

expression for probeset 5728701, but only in the ileum was significant differential

relative expression detected between control (red line) and lactation (blue line).
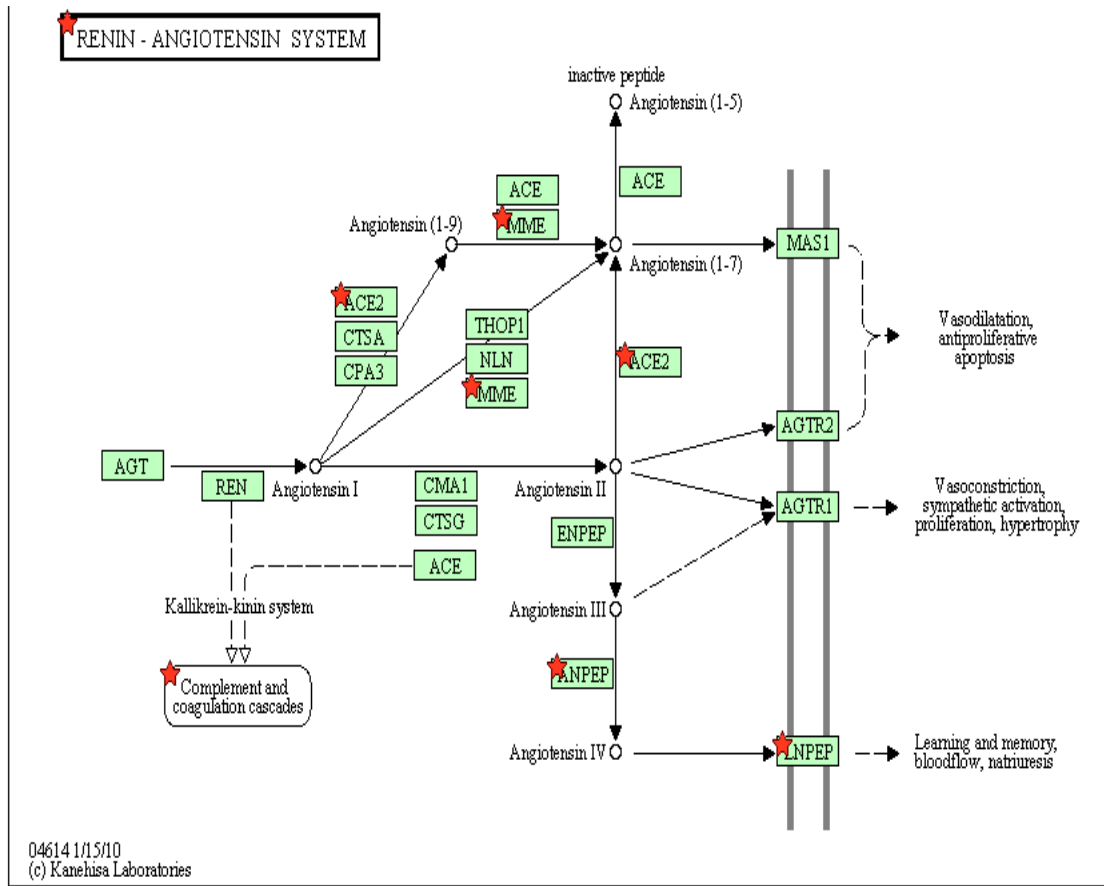
Figure 4.7: The renin angiotensin system from KEGG [126, 160, 161] as displayed in DAVID. Stars denote specific enzymes detected as differentially alternatively spliced within the pathway. Using the list of genes differentially alternatively spliced in the ileum, the renin angiotensin pathway displays a p-value of 5.4×10^{-4}.

123

Tables

Table 4-1: Genes detected to be differentially spliced in the liver of lactating animals

Gene Symbol	# of probesets	Transcript Cluster ID	RefSeq	alt splicing p
Bhmt	7	7202449	NM_030850	1.07E-08
Cdc42	8	7292247	AF205635	1.79E-08
Cpa1	9	7252157	NM_016998	1.07E-07
Gck	13	7127894	NM_012565	1.55E-07
Lss	26	7221620	NM_031049	1.63E-06
Ugt1a	23	7357340	NM_201425	1.19E-05
Slc6a13	14	7258065	NM_133623	2.59E-05
Comt	6	7089781	NM_012531	3.64E-05

ANOSVA was utilized to identify genes that were likely differentially spliced with a cutoff of an FDR of 0.05. Bhmt: Betaine-homocysteine methyltransferase; Cdc42: cell division cycle 42; Cpa1: Carboxypeptidase A1; Gck: glucokinase; Lss:

Lanosterol synthase; Ugt1a: UDP glucuronosyl transferase 1 family; Slc6a13: Solute carrier family 6 member 13; Comt: Catechol-O-methyltransferase. "Gene symbol" is the NCBI official gene symbol for the respective transcript, "# of probesets" indicates the number of probesets associated with the transcript, "Transcript Cluster ID" indicates the Affymetrix transcript cluster ID associated with the transcript, "RefSeq" identifies the Refseq entry for the corresponding transcript, and "alt splicing p" indicates the interaction p-value between probeset ID and physiologic state effect given by Partek.

Table 4-2: Genes were detected to be differentially spliced in the duodenum of lactating animals.

Gene Symbol	# of probesets	Transcript Cluster ID	RefSeq	alt splicing p
Reg1a	6	7255179	---	9.19E-10
Epb4.1l3	28	7358708	NM_053927	3.57E-08
Ppap2a	7	7189673	NM_022538	1.36E-06
Slc34a2	15	7126150	NM_053380	1.48E-06
Ctrl	7	7184989	NM_054009	1.77E-05
Ptprh	14	7028616	D45413	1.95E-05
Abcb9	12	7098148	NM_022238	2.87E-05
Fbxo8	8	7150879	---	4.50E-05
Atf2	12	7240317	---	7.32E-05

ANOSVA was utilized to identify genes that were likely differentially spliced with a cutoff of an FDR of 0.05. Reg1a: Regenerating islet derived alpha 1; Ebp4.1l3: Erythrocyte membrane protein band 4.1-like 3; Ppap2a: Phosphatidic acid phosphatase type 2A; Slc34a2: Solute carrier family 34 member 2; Ctrl: Chymostrypsin-like; Ptprh: Protein tyrosine phosphatase receptor type H; Abcb9: ATP-binding cassette subfamily b member 9; Fbxo8: F-box protein 8; Atf2: Activating transcription factor 2. Column headings are as used in table 4.1.

Table 4-3: Genes detected to be differentially spliced in the jejunum of

lactating animals.

Gene Symbol	# of probesets	Transcript Cluster ID	RefSeq	alt splicing p
Acta1	7	7186461	NM_019212	6.10E-09
Alpi2	15	7357124	NM_022680	3.82E-08
LMO7	31	7134690	NM_001001515	4.38E-06
Maob	16	7373289	NM_013198	2.09E-05
Ppap2a	7	7189673	NM_022538	4.01E-05
Tpm4	8	7149580	NM_012678	7.02E-05

ANOSVA was utilized to identify genes that were likely differentially spliced with a cutoff of an FDR of 0.05. Acta1: actin alpha 1, Alpi2: Alkaline phosphatase 3; LMO7: Lim domain 7; Maob: monoamine oxidase b; Ppap2a: Protein tyrosine phosphatase type 2A; Tpm4: Tropomyosin 4.   Column headings are as used in Table 4.1.

Table 4.4: Genes found in the "alternative splicing" Sp-Pir keywords category.

Transcript_Cluster_ID	Gene Name
7314252	ATPase, Ca++ transporting, plasma membrane 1
7137113	ERO1-like (S. cerevisiae)
7094780	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1
7357340	UDP glucuronosyltransferase 1 family, polypeptide A2; UDP glucuronosyltransferase 1 family, polypeptide A5; UDP glycosyltransferase 1 family polypeptide A3; UDP glucuronosyltransferase 1 family, polypeptide A6; UDP glucuronosyltransferase 1 family, polypeptide A9; UDP glycosyltransferase 1 family, polypeptide A8; UDP glucuronosyltransferase 1 family, polypeptide A7C; UDP glucuronosyltransferase 1 family, polypeptide A1
7070578	acetyl-coenzyme A carboxylase alpha
7083901	archaelysin family metallopeptidase 2
7085242	casein kinase 1, delta

Transcript_Cluster_ID	Gene Name
7292247	cell division cycle 42 (GTP binding protein)
7353844	interleukin 1 receptor-like 1
7146425	neuregulin 1
7160039	Optineurin
7033968	phosphatidylinositol binding clathrin assembly protein
7192686	phospholipase D1
7289791	sterol carrier protein 2
7091588	synaptojanin 1
7347403	transcription factor 12
7346991	tropomyosin 1, alpha

Transcript_Cluster_ID refers to the Affymetrix transcript clusters that identify each gene on the exon array, and gene name identifies the genes by name.

Additional File 4-1: <u>Differentially_spliced_genes_ileum.txt</u>. ANOSVA was utilized to identify genes that were likely differentially spliced with a cutoff of an FDR of 0.05. Column headings are as used in Table 4.1. Gene assignment is a detailed description of the full name of the gene, and contains the refseq entry and gene symbol. Given in .txt format.

Additional File 4-2: <u>Abcg8_sequence.txt</u> Sequence of both the full Abcg8 transcript and the detected splice variant. Given in .txt format.

Chapter 5

Discussion, Additional Studies, and Conclusions

Cholesterol Biosynthesis is upregulated at the mRNA level in lactating dams

Analysis of the rat exon array data revealed several changes at the mRNA level that aid in defining the physiology of the lactating rat. Fisher's exact test in Ingenuity Pathways (IPA) (Ingenuity Systems, www.ingenuity.com) identified pathways that were altered at day 10, 16h lactating rats when compared to virgin controls. The most evident change was increased mRNA from genes in the cholesterol biosynthetic pathway. Lactating rats need cholesterol both for secretion into the milk as well as to account for the increased production of bile acids [10]. 3-Hydroxy-3-methylglutaryl-coenzyme A reductase (Hmgcr), the rate limiting step of cholesterol synthesis was upregulated in the liver, duodenum, and ileum. This was consistent with early data demonstrating increased cholesterol synthesis in the liver and small intestine [9, 111], and increased Hmgcr activity in the liver during lactation [86, 112]. As the cholesterol biosynthetic genes are collectively controlled by sterol response element regulatory element binding protein 2 (Srebp-2), a very likely mechanism for increased mRNA production of the cholesterol biosynthetic enzymes was increased Srebp-2 activity.

As the cholesterol biosynthetic genes are all controlled by the transcription factor Srebp-2, the activity of Srebp-2 in hepatocytes, enterocytes, and mammary epithelial cells could prove useful in characterizing how gene expression is

131

controlled during lactation. As there is some overlap between targets of the

Srebp proteins [87], the activities of Srebp-1a and Srebp-1c should also be

investigated. In addition, Srebp-1a and Srebp-2 can form a heterodimer, and

activity of the heterodimer may also be relevant in the tissues studied [182, 183].

It has been proposed that in the lactating mammary gland, exposure to prolactin

activates Akt, which in turn results in increased amounts of nuclear Srebp-1c

[184].

Nuclear amounts of Srebp-1a, Srebp-1c and Srebp-2 can be determined by

Western analysis of nuclear extracts of the tissues of interest [185] (in this case

the hepatocyte and enterocytes from each part of the small intestine.) Srebp-1c

has also been specifically identified in the nuclear extracts of rat hepatocytes by

Western blot [186].

**Slc39a4 shows increased mRNA in the gastrointestinal tract of lactating
dams**

In addition to utilizing pathway analysis through Ingenuity Pathways, genes were

grouped according to where in the gastrointestinal tract they display differential

gene expression. The zinc transporter Slc39a4 was in the list of genes

upregulated in the liver, duodenum, jejunum, and ileum.

Increased Slc39a4 is consistent with reports of increased zinc uptake during

lactation in rats [76] and humans [73, 74]. As Slc39a4 is the major, although not

only route through which zinc is absorbed in the small intestine [132], its

increased expression during lactation provides a likely mechanism for increased zinc absorption.

Here, we hypothesize that increased expression of active Slc39a4 on the apical domain of enterocytes is responsible for the improved zinc absorption observed during lactation. First, if the increase in Slc39a4 is leading to increased absorption, then the increase is likely happening at the level of protein as well, and this should be detectable on the apical membrane of enterocytes by Western blot and immunohistochemistry [187]. Because increased mRNA levels were detected in the microarray, it is worth detecting if the relative amounts of Slc39a4 protein are consistent with the level of gene expression detected in the microarray. Under conditions of sufficient zinc concentration, Slc39a4 is rapidly degraded. Therefore, it would be important to analyze levels of the protein throughout the day in order to detect if and when protein levels are altered. In addition, because intracellular zinc concentration regulates stabilization of Slc39a4 mRNA and protein, dietary intake of could be monitored in order to compare presence of Slc39a4 protein relative to the amount of zinc consumed in the diet.

The ileum expresses increased differential alternative splicing in lactating animals compared to the liver, duodenum, and jejunum

Utilization of a two-way ANOVA for probeset effect and physiologic state effect in Partek (Partek Incorporated, St. Louis, Missouri) resulted in detection of a much higher number of positive tests for the interaction term in the ileum (89) than in

the liver, duodenum, and jejunum (less than 20 in each tissue). A positive test was indicative of differential alternative splicing, suggesting that regulation of differential splicing differs in the ileum relative to the other tissues. Currently, we do not have an explanation regarding what could be regulating differential splicing in the ileum.

The models for Analysis of Splicing Variance (ANOSVA) and Microarray Detection of Differential Alternative Splicing (MiDAS) overestimate p-values

Histograms of p-values for both ANOSVA in Partek and MiDAS indicated a bias toward p-values at one (1), rather than p-values being uniformly distributed when not significant. This bias has been indicated in the literature [109] and suggested that the models do not accurately reflect the data. We propose the following workflow to evaluate the current existing methods for differential analysis of alternative splicing.

Considering that both linear models (ANOSVA and MiDAS) appear to overestimate p-values, as shown by our own data and reported elsewhere [97, 109], perhaps a parametric statistical model is not optimal for identifying locations of alternative splicing. Regarding non-parametric tests, a Rank Product approach has been applied, but it is computationally demanding at the gene level [109]. In addition, although corporate software such as JMP Genomics and Partek allow for some filtering steps to be readily applied, such as specifying a metaprobeset file which defines which of Affymetrix's confidence levels are used and setting a minimal intensity filter, other filtering steps need to be applied

manually. For example, Affymetrix suggests that probesets with intensities that are higher than the constitutively expressed exons are likely cross-hybridizing with cDNA other than the expected target [97]. Validation rates through RT-PCR in published literature range from less than 50% [188] to above 80% [110] . This clearly indicates the need for effective filtering to remove genes that appear to test positive for various reasons, but are not differentially alternatively spliced. Note that the phrase "differentially alternatively spliced" is used loosely, as calculations are based on a measurement of exon intensity relative to gene intensity. Consequently, events such as alternative transcription start sites and polyadenylation site usage will also be detected as "alternative splicing".

A front end package that allows the user to readily apply various filtering algorithms as well as run a variety of statistical measures for differential alternative splicing would prove very useful. We propose the following workflow to integrate the existing methods for detection of alternative splicing into one package (Figure 5.1). It has been proposed that utilizing multiple methods and selecting the overlapping region of significant genes reduces the risk of false positives [109], so providing a package that reports the results from these tests and is able to generate a Venn Diagram of the overlapping regions of different tests could prove useful. Although software such as JMP and Partek exist, these software packages tend to rely on an ANOVA model to detect differences in differential alternative splicing.

The first step to analyzing the microarray alternative splicing data is allowing the software to access the data. There are two possible situations in which an investigator may wish to import the data. In the first case, the investigator has raw files directly following analysis of a dataset. In the second case, the investigator has already converted the data into a workable format (for example, by running RMA) and simply wishes to import the data as a spreadsheet. In the first case, various options for background correction, normalization, and summarization should be available. Note that the option to not summarize the data, but rather work with the probe level data should be available in case the experimenter wishes to work directly from this data [189]. It is important that we include options for RMA when importing (RMA background correction, quantile normalization, and median polish) although other options may be included such as median normalization, and Tukey's biweight for summarization. Note that separate code would probably need to be written for each platform when importing raw data files, as each company uses its own file format. Consequently, it may be simpler to assume that the experimenter has the data already transformed and ready to be analyzed in a .txt format, in which case the data can be input as a large matrix. In order to perform all of the previously described algorithms, data need to be available at the probe level, probeset (or exon) level, and gene level, meaning three datasets ultimately need to be input, in addition to a file that relates the multiple datasets. For Affymetrix datasets, this is the cel design file (cdf), but a table with three columns (probe identifier, probeset (or exon) identifier, and gene identifier) should be sufficient and readily

available with a small amount of work for most experimenters on most platforms. Also, it should be worthwhile to note that information other than expression measurements may wish to be imported, such as DABG p-values for Affymetrix datasets for the purposes of filtering out poorly expressed probesets. When importing, investigators should be asked which values are associated with microarrays, which microarrays are associated with respective treatment groups, and which columns include other data.

Filtering data is an essential step in removing false positives, as nearly all methods for detecting differential splicing events run into similar problems with genes that are poorly expressed in a single group, exons that are poorly expressed in all groups, and probes that cross-hybridize. Affymetrix has a series of recommended filtering procedures [97], but finding software that readily applies all of the available filters is difficult. Through spreadsheet manipulation, the filters can be applied in JMP Genomics (SAS Institute Inc., Cary NC) or Excel (Microsoft Corporation), but a streamlined approach has not been taken for many of the possible filters. Filtering out genes that are poorly expressed in either group, exons that are poorly expressed in all groups, and probesets that are likely to cross-hybridize are all essential [97]. Presence/absence calls can be made based on a value in a column being greater than or less than a specified number. For example, one could make a first set of presence/absence calls under the conditions where if intensity for a given exon on a chip>3, then that exon on that chip is present; otherwise the exon is absent. We could do the

same for DABG values (if DABG<0.05 then present, otherwise is absent). The "cross-hyb type" column in the Affymetrix annotation files indicates whether or not a probe is predicted to cross-hybridize with non-target cDNA. If cross-hyb type does not equal one (which is defined as not likely to cross-hybridize to non-target cDNA), then an absence call can be made for that probe on all chips, indicating that the respective exon is to be filtered out. Affymetrix's recommendation for applying a presence/absence call for gene level data is whether 50% of the exons are present, so a presence/absence call could be made from the probeset level data accordingly. Under normal circumstances, a gene should be removed if it is not expressed in any group, and a probeset/exon should be removed if none of the groups express the probeset/exon.

Both JMP Genomics and Partek rely on a two-way ANOVA method for detection of differential splicing analogous to the ANOSVA method. However, other methods exist including MiDAS [97], FIRMA [108], PECA-SI [189], and analysis of the rank products of splicing indices or intensities [109]. R-code exists for these methods, and given the open-source nature of R, these codes could be edited with credit to the original authors [108, 109, 189]. Ideally, software would ask the investigator which method should be applied and then analyze the corresponding method. Like many other steps, clever spreadsheet manipulation makes these methods technically available using other software, although code to run these methods has not been specifically written for the spreadsheet management software. However, organizing the methods so that they are

readily executable would be a useful tool. Following analyses, spreadsheet software or the proposed software should be able to generate a Venn diagram of genes that test positive at various cutoffs (either raw p-values or FDR correcting methods can be used) from the various output files. Note that many of these steps are computationally intensive, so managing how memory is used could become an issue. Ideally, this should give a mechanism for readily comparing the large number of methods for detecting alternative splicing that exist. The overall workflow can be found in Figure 5.1.

An alternative transcript was detected for Abcg8 in the lactating liver

Abcg8, part of a heterodimer that transports cholesterol into bile, shows decreased mRNA in the liver of lactating rats [93, 164], and functional Abcg8 protein is not detectable in the hepatocytes of lactating dams [93]. In preliminary analysis of the alternative splicing dataset, Abcg8 was identified as a likely candidate for differential alternative splicing in the liver of lactating dams. We therefore ran RT-PCR on the liver samples and isolated a band that was unique to a clone in lactating liver. Sequencing revealed that a region that begins in part of exon four and ends in part of exon nine was spliced out in the splice variant. This further indicated part of the control of Abcg5/Abcg8 activity in the liver in lactating rats is controlled at the mRNA level.

Possible differential alternative splicing in the renin-angiotensin system was detected in the ileum

Analysis of the list of genes that tested positive in the ileum by DAVID [105] resulted in identification of the renin-angiotensin as being a candidate for differential regulation in lactation through alternative splicing. Of the four genes detected to be differentially alternatively spliced in Partek, one, angiotensin converting enzyme 2 (Ace2) had the lowest alternative splicing p-value amongst any gene in the dataset. Ace2 is important in inhibiting the renin-angiotensin pathway, so validating the presence of the detected splice variant, and characterizing if it has altered function would be useful in describing the renin-angiotensin pathway in the lactating rat.

Future work: Characterization of Differences in Gene Expression in the liver Throughout Pregnancy and Lactation

Our present work identified differences in gene expression in the liver and small intestine of lactating rats compared to age matched virgin controls at day 10 postpartum, 16 h. This only characterizes the difference in gene expression relative to controls at one specific time point. One important question would be how the changes identified in lactating rats at this time point differ from changes at other time points. For example, our laboratory has identified that day 10, 16h is the first point at which Cyp7a1, the rate limiting step for bile acid synthesis, shows a statistically significant increase in mRNA expression [10]. However, we have noted a number of physiologic differences regarding expression of hepatic bile acid transporter in early (day 2 postpartum) lactation [82]. Taking samples across several stages in lactation and pregnancy would allow an investigation for

when changes in gene expression begin and end. If trends in gene expression are associated with specific pathways or transcription factors, this would also allow characterization of the roles of these pathways during lactation. We propose measuring gene expression through microarray at the following time points: virgin (16 h), early pregnancy, late pregnancy, early lactation (day 2 postpartum, 16 h), middle lactation (day 10 postpartum, 16h), and late lactation. Following the previous pooling scheme of four replicates per group and four individual rats composing a single replicate may lead to an excessive number of required rats (4 rats X 4 replicates X 6 groups = 96 rats). The absence of a pooling scheme would result in higher variance amongst individuals, meaning that the statistical power would be based solely on the four replicates. A possible compromise would be to reduce the number of rats pooled in a sample to three, requiring a total of 72 rats for the entire experiment. We chose to focus on the liver because of its role in regulating bile acid synthesis, metabolizing xenobiotics, as well as being a major site of cholesterol biosynthesis. Similar experiments have been performed in the mammary gland of pregnant and lactating mice [184, 190]. If we choose to only analyze a specific pathway, such as the cholesterol biosynthetic pathway, a small custom chip or nanostring experiment would be more efficient, and statistical power would be improved by reducing the number of genes queried.

A one-way ANOVA model would be used for the purposes of determining statistical significance of each gene, with time being treated as a class variable.

The overall ANOVA would initially be fitted to an FDR of 0.2. Genes that pass

the initial test will need to be tested for pairwise comparisons between the

respective time points. To simplify comparisons, we would only consider the

differences of each time point relative to the virgin controls. This reduces the

number of pairwise comparisons performed for each positive testing gene from

15 to 5. We currently propose a significance cutoff of $p<0.0020$ for the statistical

contrasts, but both this value and the FDR may need to be adjusted based on the

size of the list of significant genes. (The p-value for contrasts within a gene is

based on a modification of the Bonferroni procedure at $\alpha=0.01$, as

$0.01/5=0.0020$). If the list is not sufficiently large, then pathway analysis

becomes difficult. Statistical pattern matching could then be used to assign

genes to groups. Each time point other than the virgin time point could be

assigned to increased, no detectable change, or decreased relative to virgins for

each gene. For the purposes of pattern matching, the pairwise comparison p

value cutoff may need to be relaxed to $p<0.05$, but only if the gene was

previously defined as differentially expressed. This is caused by the fact that a

false negative would lead to assignment of a gene to a separate group.

Characterization of these groups should give an indication as to when genes

increase and decrease expression during pregnancy and lactation, and could

lead to identification of signaling pathways that mediate changes in gene

expression.

Conclusion

We showed through microarray studies increased mRNA of cholesterol biosynthetic genes in the liver and small intestine and a likely mediator of increased zinc uptake (Slc39a4). We propose that increased Srebp activity, particularly Srebp-2 is involved in the increased cholesterol synthesis detected in the liver and small intestine of lactating dams. In addition, we propose that the difference in expression in Insig in the small intestine and Scap in the liver causes the more dramatic increase in cholesterol synthesis in the liver relative to the small intestine. Further work needs to be done to elucidate the mechanisms regarding how these changes in mRNA influence the physiology of the lactating animal. In addition, we have proposed candidates for differential alternative splicing between lactating animals and virgin controls and have detected a splice variant of Abcg8 in the livers of lactating animals. We also identified a possible splice variant for Ace2 in the ileum that requires verification. Depending on the function of the splice variant, the Ace2 variant may influence sodium absorption and blood pressure in the lactating animal. A number of additional studies can be performed, both to further characterize alternative splicing in more detail, and to elucidate how the detected changes in gene expression and alternative splicing influence the physiology of the lactating animal.
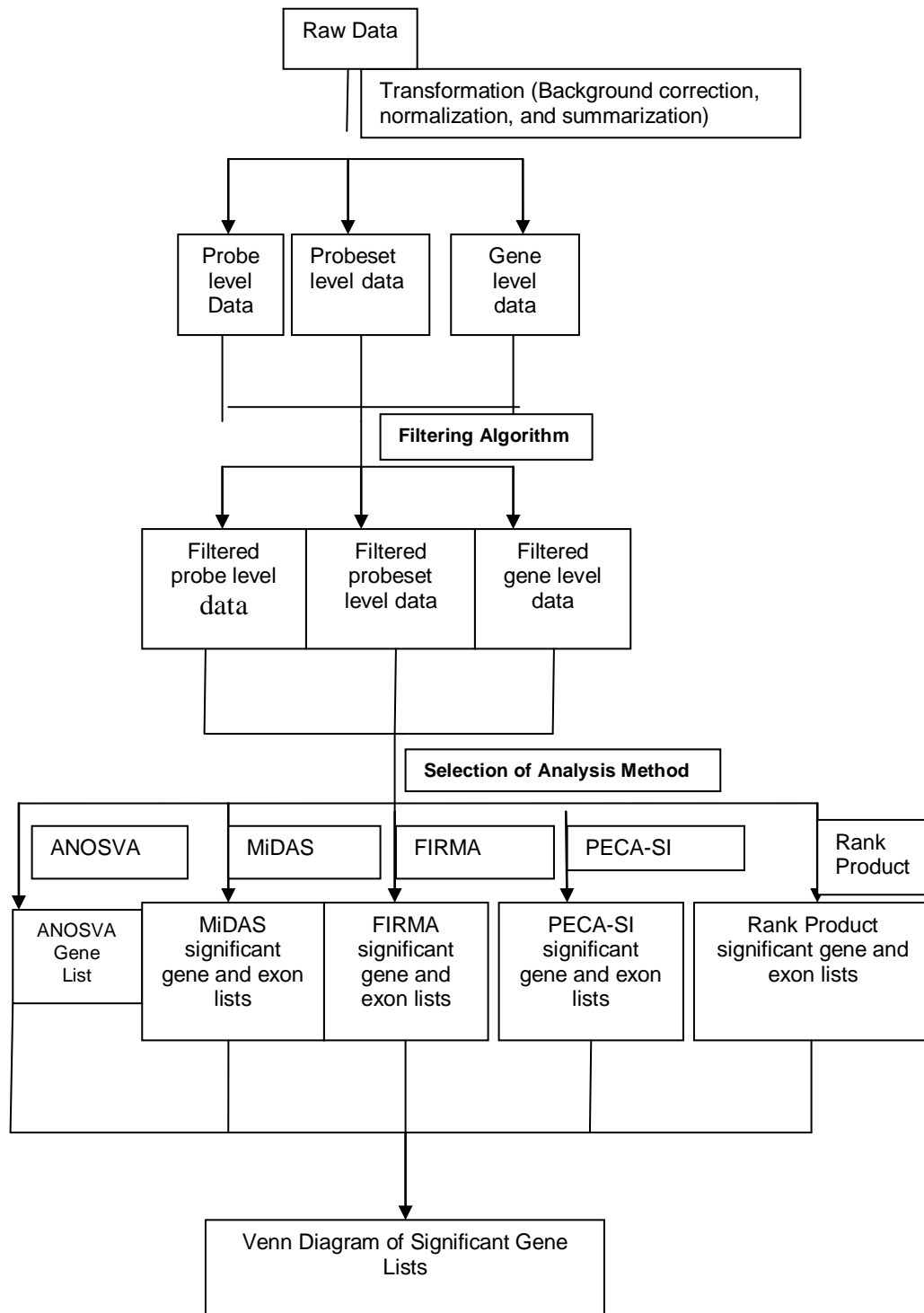
Figure 5.1: Workflow for detection of differential alternative splicing. Data are initially imported and saved as gene level, probeset level (exon level), and/or at the level of individual probes, depending on the summarization method used. From here, data that would likely contribute to false positives are filtered out. Following the filtering algorithm, various methods can be applied to analyze the datasets, calling up filtered data at the appropriate summarization levels (gene, probeset, or probe) as input. Venn diagrams can be used to compare results across the various methods by cross-referencing which genes, and in some cases exons, test positive by the various tests.

Appendix A: List of Abbreviations

Prl-Prolactin

Abc-Adenosine Triphosphate Binding Cassette

Thra- Thyroid receptor alpha

Thrb-Thyroid receptor beta

Klf9-Kruppel like factor 9

TRα1-Thyroid receptor alpha isoform 1

TRβ1-Thryroid receptor beta isoform 1

T3-triiodothyrodine

PI3K-phosphoinositide 3 kinase

AKT- v-akt murine thymoma viral oncogene homolog

PIP2-Phosphotidyl inositol 4,5 bisphosphate

PTEN-Phosphotase and tensin homolog

PDK1-Pyruvate dehydrogenase kinase 1

FOXO-Forkhead box transcription factor

mTOR- Mechanistic target of rapamycin

4EBP1-Eukaryotic initiation factor 4E binding protein 1

PEPCK-Phosphoenolpyruvate carboxykinase

G6P-Glucose 6 Phosphotase

AMPK-Adenosine monophophate kinase

Hmgcr-3-Hydroxy-3-methylglutaryl Coenzyme A reductase

Glut4- Glucose transporter 4

Acc-Acetyl-Coa Carboxylase

CRTC2-CREB regulated transcription coactivator 2

SREBP-Sterol regulatory element binding protein

Ppargc1α-Ppar gamma coactivator 1α

JAK-Janus Kinase

STAT-Signal transducer and activator of transcription

Prlr-Prolactin receptor

SR-Serine arginine

Slc-Solute carrier

Cyp-Cytochrome P450

Erk-Extracellular signal related kinase

Asbt- Apical sodium dependent bile acid transporter

Ntcp- Sodium taurocholate cotransporting polypeptide

Srebf- Sterol regulatory element binding factor

Scap-SREBP chaperone

Ugt- UDP glucuronosyltransferase

DNA-Deoxyribonucleic acid

RNA-Ribonucleic Acid

RMA-Robust multichip averaging

ANOVA-Analysis of variance

FWER-Family wise error rate

FDR-False Discovery Rate

PFP-Proportion of false positives

DAVID-Database for annotation visualization and integrated discovery

EST-Expressed sequence tag

MiDAS-Microarray detection of alternative splicing

FIRMA-Finding isoforms through robust multichip averaging

ANOSVA-Analysis of splicing variance

DABG-Detection above background

Ace2-Angiotensin converting enzyme 2

Ostα/β-Organic solute transporter alpha/beta

Bsep-Bile salt export pump

FGF-Fibroblast growth factor

Fgfr-Fibroblast growth factor receptor

RT-PCR-Reverse transcriptase polymerase chain reaction

Rg9-mtd2- RNA guanine-9 methyltransferase domain containing 2

KEGG-Kyoto Encyclopedia of genes and genomes

Ch25h-Cholesterol-25 hydrolase

Akr1d1-Aldo keto reducatase family 1 member d1

Acox2-Acetyl-CoA oxidase 2

Scp2-Sterol carrier protein 2

Baat-Bile acid-CoA:amino acid N-acyltransferase

IPA-Ingenuity Pathways Analysis

CD28-Cluster of differentiation 28

IL-2- Interleukin 2

Insig-Insulin induced gene

Scd2-Stearoyl-Coenzyme A desaturase 2

ME1- Malic enzyme 1

Ctsb-Cathepsin B

Tmbim6-Transmembrane BAX inhibitor motif containing 6

Tmed2-Transmembrane emp24 domain trafficking protein 2

Tmem97-Transmembrane protein 97

NPC1l1-Niemann Pick 1 like 1

Fdft1-Farnesyl-diphosphate farnesyl transferase 1

Mvk-Mevalonate Kinase

Pmvk-Phosphomevalonate kinase

Mvd-Mevalonate decarboxylase

Idi1-Isopentyl-diphosphate delta isomerase 1

Sqle-Squalene epoxidase

Lss-Lanosterol synthase

Dhcr7-7-dehydro cholesterol reductasse

BLAT-BLAST like alignment tool

UCSC-University of California, Santa Cruz

Anpep-Alanyl aminopeptidase

Lnpep-Leucyl/cystinyl peptidase

Mme-Membrane metallo-endopeptidase

Ahr-Aromatic hydrocarbon recptor

References

1. Hammond KA: **Adaptation of the maternal intestine during lactation**. *J Mammary Gland Biol Neoplasia* 1997, **2**(3):243-252.
2. Cripps AW, Williams VJ: **The effect of pregnancy and lactation on food intake, gastrointestinal anatomy and the absorptive capacity of the small intestine in the albino rat**. *Br J Nutr* 1975, **33**(1):17-32.
3. Vernon RG, Denis RG, Sorensen A, Williams G: **Leptin and the adaptations of lactation in rodents and ruminants**. *Horm Metab Res* 2002, **34**(11-12):678-685.
4. Ben-Jonathan N, LaPensee CR, LaPensee EW: **What can we learn from rodents about prolactin in humans?** *Endocr Rev* 2008, **29**(1):1-41.
5. Delouis C, Dijiane J, Houdebine LM, Terqui M: **Relation between hormones and mammary gland function**. *J Dairy Sci* 1980, **63**(9):1492-1513.
6. Knight CH, Peaker M: **Development of the mammary gland**. *J Reprod Fertil* 1982, **65**(2):521-536.
7. Neville MC, McFadden TB, Forsyth I: **Hormonal regulation of mammary differentiation and milk secretion**. *J Mammary Gland Biol Neoplasia* 2002, **7**(1):49-66.
8. Liu Y, Hyde JF, Vore M: **Prolactin regulates maternal bile secretory function post partum**. *J Pharmacol Exp Ther* 1992, **261**(2):560-566.
9. Feingold KR, Moser AH: **Effect of lactation on cholesterol synthesis in rats**. *Am J Physiol* 1985, **249**(2 Pt 1):G203-208.
10. Wooton-Kee CR, Cohen DE, Vore M: **Increased cholesterol 7alpha-hydroxylase expression and size of the bile acid pool in the lactating rat**. *Am J Physiol Gastrointest Liver Physiol* 2008, **294**(4):G1009-1016.
11. Baldwin RL, McLeod KR, Capuco AV: **Visceral tissue growth and proliferation during the bovine lactation cycle**. *J Dairy Sci* 2004, **87**(9):2977-2986.
12. Viturro E, Koenning M, Kroemer A, Schlamberger G, Wiedemann S, Kaske M, Meyer HH: **Cholesterol synthesis in the lactating cow: Induced expression of candidate genes**. *J Steroid Biochem Mol Biol* 2009, **115**(1-2):62-67.
13. Langley LLT, Ira R.; Christensen. John B.: **Dynamic Anatomy and Physiology** New York, NY: McGraw-Hill Book Company; 1974.
14. Rost D, Mahner S, Sugiyama Y, Stremmel W: **Expression and localization of the multidrug resistance-associated protein 3 in rat small and large intestine**. *Am J Physiol Gastrointest Liver Physiol* 2002, **282**(4):G720-726.
15. Oostendorp RL, Beijnen JH, Schellens JH: **The biological and clinical role of drug transporters at the intestinal barrier**. *Cancer Treat Rev* 2009, **35**(2):137-147.
16. Zimmermann C, Gutmann H, Hruz P, Gutzwiller JP, Beglinger C, Drewe J: **Mapping of multidrug resistance gene 1 and multidrug resistance-associated protein isoform 1 to 5 mRNA expression along the human intestinal tract**. *Drug Metab Dispos* 2005, **33**(2):219-224.

17. Dietrich CG, Geier A, Oude Elferink RP: **ABC of oral bioavailability: transporters as gatekeepers in the gut**. *Gut* 2003, **52**(12):1788-1795.

18. Cheng SY, Leonard JL, Davis PJ: **Molecular aspects of thyroid hormone actions**. *Endocr Rev* 2010, **31**(2):139-170.

19. Denis RG, Williams G, Vernon RG: **Regulation of serum leptin and its role in the hyperphagia of lactation in the rat**. *J Endocrinol* 2003, **176**(2):193-203.

20. Capuco AV, Kahl S, Jack LJ, Bishop JO, Wallace H: **Prolactin and growth hormone stimulation of lactation in mice requires thyroid hormones**. *Proc Soc Exp Biol Med* 1999, **221**(4):345-351.

21. Quevedo-Corona L, Franco-Colin M, Caudillo-Romero M, Pacheco-Rosado J, Zamudio-Hernandez S, Racotta R: **3,5,3'-Triiodothyronine administered to rat dams during lactation increases milk yield and triglyceride concentration and hastens pups growth**. *Life Sci* 2000, **66**(21):2013-2021.

22. Aceves C, Valverde C: **Type I, 5'-monodeiodinase activity in the lactating mammary gland**. *Endocrinology* 1989, **124**(6):2818-2820.

23. Keijzer R, Blommaart PJ, Labruyere WT, Vermeulen JL, Doulabi BZ, Bakker O, Tibboel D, Lamers WH: **Expression of thyroid hormone receptors A and B in developing rat tissues; evidence for extensive posttranscriptional regulation**. *J Mol Endocrinol* 2007, **38**(5):523-535.

24. Moeller LC, Dumitrescu AM, Walker RL, Meltzer PS, Refetoff S: **Thyroid hormone responsive genes in cultured human fibroblasts**. *J Clin Endocrinol Metab* 2005, **90**(2):936-943.

25. Poguet AL, Legrand C, Feng X, Yen PM, Meltzer P, Samarut J, Flamant F: **Microarray analysis of knockout mice identifies cyclin D2 as a possible mediator for the action of thyroid hormone during the postnatal development of the cerebellum**. *Dev Biol* 2003, **254**(2):188-199.

26. Feng X, Jiang Y, Meltzer P, Yen PM: **Thyroid hormone regulation of hepatic genes in vivo detected by complementary DNA microarray**. *Mol Endocrinol* 2000, **14**(7):947-955.

27. Yen PM, Feng X, Flamant F, Chen Y, Walker RL, Weiss RE, Chassande O, Samarut J, Refetoff S, Meltzer PS: **Effects of ligand and thyroid hormone receptor isoforms on hepatic gene expression profiles of thyroid hormone receptor knockout mice**. *EMBO Rep* 2003, **4**(6):581-587.

28. Miller LD, McPhie P, Suzuki H, Kato Y, Liu ET, Cheng SY: **Multi-tissue gene-expression analysis in a mouse model of thyroid hormone resistance**. *Genome Biol* 2004, **5**(5):R31.

29. Flores-Morales A, Gullberg H, Fernandez L, Stahlberg N, Lee NH, Vennstrom B, Norstedt G: **Patterns of liver gene expression governed by TRbeta**. *Mol Endocrinol* 2002, **16**(6):1257-1268.

30. Harvey CB, Stevens DA, Williams AJ, Jackson DJ, O'Shea P, Williams GR: **Analysis of thyroid hormone responsive gene expression in osteoblastic cells**. *Mol Cell Endocrinol* 2003, **213**(1):87-97.

31. Weitzel JM, Hamann S, Jauk M, Lacey M, Filbry A, Radtke C, Iwen KA, Kutz S, Harneit A, Lizardi PM *et al*: **Hepatic gene expression patterns in thyroid hormone-treated hypothyroid rats**. *J Mol Endocrinol* 2003, **31**(2):291-303.

32.   Simmen FA, Xiao R, Velarde MC, Nicholson RD, Bowman MT, Fujii-Kuriyama Y, Oh SP, Simmen RC: **Dysregulation of intestinal crypt cell proliferation and villus cell migration in mice lacking Kruppel-like factor 9**. *Am J Physiol Gastrointest Liver Physiol* 2007, **292**(6):G1757-1769.

33.   McConnell BB, Yang VW: **Mammalian Kruppel-like factors in health and diseases**. *Physiol Rev* 2010, **90**(4):1337-1381.

34.   Wrutniak C, Cassar-Malek I, Marchal S, Rascle A, Heusser S, Keller JM, Flechon J, Dauca M, Samarut J, Ghysdael J *et al*: **A 43-kDa protein related to c-Erb A alpha 1 is located in the mitochondrial matrix of rat liver**. *J Biol Chem* 1995, **270**(27):16347-16354.

35.   Casas F, Rochard P, Rodier A, Cassar-Malek I, Marchal-Victorion S, Wiesner RJ, Cabello G, Wrutniak C: **A variant form of the nuclear triiodothyronine receptor c-ErbAalpha1 plays a direct role in regulation of mitochondrial RNA synthesis**. *Mol Cell Biol* 1999, **19**(12):7913-7924.

36.   Andersson ML, Vennstrom B: **Chicken thyroid hormone receptor alpha requires the N-terminal amino acids for exclusive nuclear localization**. *FEBS Lett* 1997, **416**(3):291-296.

37.   Furuya F, Lu C, Guigon CJ, Cheng SY: **Nongenomic activation of phosphatidylinositol 3-kinase signaling by thyroid hormone receptors**. *Steroids* 2009, **74**(7):628-634.

38.   Hiroi Y, Kim HH, Ying H, Furuya F, Huang Z, Simoncini T, Noma K, Ueki K, Nguyen NH, Scanlan TS *et al*: **Rapid nongenomic actions of thyroid hormone**. *Proc Natl Acad Sci U S A* 2006, **103**(38):14104-14109.

39.   Castaneda CA, Cortes-Funes H, Gomez HL, Ciruelos EM: **The phosphatidyl inositol 3-kinase/AKT signaling pathway in breast cancer**. *Cancer Metastasis Rev* 2010, **29**(4):751-759.

40.   Iglesias PA: **Spatial regulation of PI3K signaling during chemotaxis**. *Wiley Interdiscip Rev Syst Biol Med* 2009, **1**(2):247-253.

41.   Zoncu R, Efeyan A, Sabatini DM: **mTOR: from growth signal integration to cancer, diabetes and ageing**. *Nat Rev Mol Cell Biol* 2011, **12**(1):21-35.

42.   Kops GJ, de Ruiter ND, De Vries-Smits AM, Powell DR, Bos JL, Burgering BM: **Direct control of the Forkhead transcription factor AFX by protein kinase B**. *Nature* 1999, **398**(6728):630-634.

43.   Nakae J, Kitamura T, Silver DL, Accili D: **The forkhead transcription factor Foxo1 (Fkhr) confers insulin sensitivity onto glucose-6-phosphatase expression**. *J Clin Invest* 2001, **108**(9):1359-1367.

44.   Ronnebaum SM, Patterson C: **The FoxO family in cardiac function and dysfunction**. *Annu Rev Physiol* 2010, **72**:81-94.

45.   Schwartz MW, Seeley RJ, Campfield LA, Burn P, Baskin DG: **Identification of targets of leptin action in rat hypothalamus**. *J Clin Invest* 1996, **98**(5):1101-1106.

46.   Ahima RS, Qi Y, Singhal NS, Jackson MB, Scherer PE: **Brain adipocytokine action and metabolic regulation**. *Diabetes* 2006, **55 Suppl 2**:S145-154.

47.   Zhang BB, Zhou G, Li C: **AMPK: an emerging drug target for diabetes and the metabolic syndrome**. *Cell Metab* 2009, **9**(5):407-416.

48. Canto C, Auwerx J: **AMP-activated protein kinase and its downstream transcriptional pathways**. *Cell Mol Life Sci* 2010, **67**(20):3407-3423.

49. Hardie DG: **AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy**. *Nat Rev Mol Cell Biol* 2007, **8**(10):774-785.

50. Shaw RJ, Lamia KA, Vasquez D, Koo SH, Bardeesy N, Depinho RA, Montminy M, Cantley LC: **The kinase LKB1 mediates glucose homeostasis in liver and therapeutic effects of metformin**. *Science* 2005, **310**(5754):1642-1646.

51. Thorn SR, Ehrhardt RA, Butler WR, Quirk SM, Boisclair YR: **Insulin regulates hepatic leptin receptor expression in early lactating dairy cows**. *Am J Physiol Regul Integr Comp Physiol* 2008, **295**(5):R1455-1462.

52. Sahu A: **Leptin signaling in the hypothalamus: emphasis on energy homeostasis and leptin resistance**. *Front Neuroendocrinol* 2003, **24**(4):225-253.

53. Oakes SR, Rogers RL, Naylor MJ, Ormandy CJ: **Prolactin regulation of mammary gland development**. *J Mammary Gland Biol Neoplasia* 2008, **13**(1):13-28.

54. Chilton BS, Hewetson A: **Prolactin and growth hormone signaling**. *Curr Top Dev Biol* 2005, **68**:1-23.

55. Wallis M: **The molecular evolution of pituitary growth hormone prolactin and placental lactogen: A protein family showing variable rates of evolution**. *Journal of Molecular Evolution* 1981, **17**(1):10-18.

56. Watson CJ, Burdon TG: **Prolactin signal transduction mechanisms in the mammary gland: the role of the Jak/Stat pathway**. *Rev Reprod* 1996, **1**(1):1-5.

57. Zhao AZ, Bornfeldt KE, Beavo JA: **Leptin inhibits insulin secretion by activation of phosphodiesterase 3B**. *J Clin Invest* 1998, **102**(5):869-873.

58. Zhao AZ, Shinohara MM, Huang D, Shimizu M, Eldar-Finkelman H, Krebs EG, Beavo JA, Bornfeldt KE: **Leptin induces insulin-like signaling that antagonizes cAMP elevation by glucagon in hepatocytes**. *J Biol Chem* 2000, **275**(15):11348-11354.

59. Zhao AZ, Huan JN, Gupta S, Pal R, Sahu A: **A phosphatidylinositol 3-kinase phosphodiesterase 3B-cyclic AMP pathway in hypothalamic action of leptin on feeding**. *Nat Neurosci* 2002, **5**(8):727-728.

60. Hirsch E, Costa C, Ciraolo E: **Phosphoinositide 3-kinases as a common platform for multi-hormone signaling**. *J Endocrinol* 2007, **194**(2):243-256.

61. Berlanga JJ, Gualillo O, Buteau H, Applanat M, Kelly PA, Edery M: **Prolactin activates tyrosyl phosphorylation of insulin receptor substrate 1 and phosphatidylinositol-3-OH kinase**. *J Biol Chem* 1997, **272**(4):2050-2052.

62. Shepard PJ, Hertel KJ: **The SR protein family**. *Genome Biol* 2009, **10**(10):242.

63. Black DL: **Mechanisms of alternative pre-messenger RNA splicing**. *Annu Rev Biochem* 2003, **72**:291-336.

64. Hertel KJ, Lynch KW, Maniatis T: **Common themes in the function of transcription and splicing enhancers**. *Curr Opin Cell Biol* 1997, **9**(3):350-357.

65. Mermoud JE, Cohen P, Lamond AI: **Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing**. *Nucleic Acids Res* 1992, **20**(20):5263-5269.

66. Mermoud JE, Cohen PT, Lamond AI: **Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism**. *EMBO J* 1994, **13**(23):5679-5688.

67. Patel NA, Chalfant CE, Watson JE, Wyatt JR, Dean NM, Eichler DC, Cooper DR: **Insulin regulates alternative splicing of protein kinase C beta II through a phosphatidylinositol 3-kinase-dependent pathway involving the nuclear serine/arginine-rich splicing factor, SRp40, in skeletal muscle cells**. *J Biol Chem* 2001, **276**(25):22648-22654.

68. Secondo A, Sirabella R, Formisano L, D'Alessio A, Castaldo P, Amoroso S, Ingleton P, Di Renzo G, Annunziato L: **Involvement of PI3'-K, mitogen-activated protein kinase and protein kinase B in the up-regulation of the expression of nNOSalpha and nNOSbeta splicing variants induced by PRL-receptor activation in GH3 cells**. *J Neurochem* 2003, **84**(6):1367-1377.

69. Vernon RG: **Lipid metabolism during lactation: a review of adipose tissue-liver interactions and the development of fatty liver**. *J Dairy Res* 2005, **72**(4):460-469.

70. Shetty PS: **Physiological mechanisms in the adaptive response of metabolic rates to energy restriction**. *Nutr Res Rev* 1990, **3**(1):49-74.

71. Mainoya JR: **Possible influence of prolactin on intestinal hypertrophy in pregnant and lactating rats**. *Experientia* 1978, **34**(9):1230-1231.

72. Muller E, Dowling RH: **Prolactin and the small intestine. Effect of hyperprolactinaemia on mucosal structure in the rat**. *Gut* 1981, **22**(7):558-565.

73. Donangelo CM, Zapata CL, Woodhouse LR, Shames DM, Mukherjea R, King JC: **Zinc absorption and kinetics during pregnancy and lactation in Brazilian women**. *Am J Clin Nutr* 2005, **82**(1):118-124.

74. Fung EB, Ritchie LD, Woodhouse LR, Roehl R, King JC: **Zinc absorption in women during pregnancy and lactation: a longitudinal study**. *Am J Clin Nutr* 1997, **66**(1):80-88.

75. Prentice A: **Micronutrients and the bone mineral content of the mother, fetus and newborn**. *J Nutr* 2003, **133**(5 Suppl 2):1693S-1699S.

76. Davies NT, Williams RB: **The effect of pregnancy and lactation on the absorption of zinc and lysine by the rat duodenum in situ**. *Br J Nutr* 1977, **38**(3):417-423.

77. Charoenphandhu N, Nakkrasae LI, Kraidith K, Teerapornpuntakit J, Thongchote K, Thongon N, Krishnamra N: **Two-step stimulation of intestinal Ca(2+) absorption during lactation by long-term prolactin exposure and suckling-induced prolactin surge**. *Am J Physiol Endocrinol Metab* 2009, **297**(3):E609-619.

78. Eide DJ: **The SLC39 family of metal ion transporters**. *Pflugers Arch* 2004, **447**(5):796-800.

79. Costello LC, Liu Y, Zou J, Franklin RB: **Evidence for a zinc uptake transporter in human prostate cancer cells which is regulated by prolactin and testosterone**. *J Biol Chem* 1999, **274**(25):17499-17504.

80. Wooton-Kee CR, Coy DJ, Athippozhy AT, Zhao T, Jones BR, Vore M: **Mechanisms for increased expression of cholesterol 7alpha-hydroxylase (Cyp7a1) in lactating rats**. *Hepatology* 2010, **51**(1):277-285.

81. Song KH, Li T, Owsley E, Strom S, Chiang JY: **Bile acids activate fibroblast growth factor 19 signaling in human hepatocytes to inhibit cholesterol 7alpha-hydroxylase gene expression**. *Hepatology* 2009, **49**(1):297-305.

82. Cao J, Huang L, Liu Y, Hoffman T, Stieger B, Meier PJ, Vore M: **Differential regulation of hepatic bile salt and organic anion transporters in pregnant and postpartum rats and the role of prolactin**. *Hepatology* 2001, **33**(1):140-147.

83. Ganguly TC, O'Brien ML, Karpen SJ, Hyde JF, Suchy FJ, Vore M: **Regulation of the rat liver sodium-dependent bile acid cotransporter gene by prolactin. Mediation of transcriptional activation by Stat5**. *J Clin Invest* 1997, **99**(12):2906-2914.

84. Mottino AD, Hoffman T, Dawson PA, Luquita MG, Monti JA, Sanchez Pozzi EJ, Catania VA, Cao J, Vore M: **Increased expression of ileal apical sodium-dependent bile acid transporter in postpartum rats**. *Am J Physiol Gastrointest Liver Physiol* 2002, **282**(1):G41-50.

85. Klaassen CD, Strom SC: **Comparison of biliary excretory function and bile composition in male, female, and lactating female rats**. *Drug Metab Dispos* 1978, **6**(2):120-124.

86. Gibbons GF, Pullinger CR, Munday MR, Williamson DH: **Regulation of cholesterol synthesis in the liver and mammary gland of the lactating rat**. *Biochem J* 1983, **212**(3):843-848.

87. Horton JD, Shah NA, Warrington JA, Anderson NN, Park SW, Brown MS, Goldstein JL: **Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes**. *Proc Natl Acad Sci U S A* 2003, **100**(21):12027-12032.

88. Horton JD, Shimomura I, Brown MS, Hammer RE, Goldstein JL, Shimano H: **Activation of cholesterol synthesis in preference to fatty acid synthesis in liver and adipose tissue of transgenic mice overproducing sterol regulatory element-binding protein-2**. *J Clin Invest* 1998, **101**(11):2331-2339.

89. Lemay DG, Neville MC, Rudolph MC, Pollard KS, German JB: **Gene regulatory networks in lactation: identification of global principles using bioinformatics**. *BMC Syst Biol* 2007, **1**:56.

90. Hylemon PB, Zhou H, Pandak WM, Ren S, Gil G, Dent P: **Bile acids as regulatory molecules**. *J Lipid Res* 2009, **50**(8):1509-1520.

91. Raghow R, Yellaturu C, Deng X, Park EA, Elam MB: **SREBPs: the crossroads of physiological and pathological lipid homeostasis**. *Trends Endocrinol Metab* 2008, **19**(2):65-73.

92. Naylor MJ, Oakes SR, Gardiner-Garden M, Harris J, Blazek K, Ho TW, Li FC, Wynick D, Walker AM, Ormandy CJ: **Transcriptional changes underlying the secretory activation phase of mammary gland development**. *Mol Endocrinol* 2005, **19**(7):1868-1883.

93.   Coy DJ, Wooton-Kee CR, Yan B, Sabeva N, Su K, Graf G, Vore M: **ABCG5/ABCG8-independent biliary cholesterol excretion in lactating rats**. *Am J Physiol Gastrointest Liver Physiol*, **299**(1):G228-235.

94.   Luquita MG, Catania VA, Pozzi EJ, Veggi LM, Hoffman T, Pellegrino JM, Ikushiro S, Emi Y, Iyanagi T, Vore M *et al*: **Molecular basis of perinatal changes in UDP-glucuronosyltransferase activity in maternal rat liver**. *J Pharmacol Exp Ther* 2001, **298**(1):49-56.

95.   Metz RP, Ritter JK: **Transcriptional activation of the UDP-glucuronosyltransferase 1A7 gene in rat liver by aryl hydrocarbon receptor ligands and oltipraz**. *J Biol Chem* 1998, **273**(10):5607-5614.

96.   Peng X, Wood CL, Blalock EM, Chen KC, Landfield PW, Stromberg AJ: **Statistical implications of pooling RNA samples for microarray experiments**. *BMC Bioinformatics* 2003, **4**:26.

97.   Affymetrix I: **Identifying and Validating Alternative Splicing Events**. In.; 2006.

98.   Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data**. *Biostatistics* 2003, **4**(2):249-264.

99.   Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data**. *Nucleic Acids Res* 2003, **31**(4):e15.

100.   Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics* 2003, **19**(2):185-193.

101.   Bretz F, Landgrebe J, Brunner E: **Multiplicity issues in microarray experiments**. *Methods Inf Med* 2005, **44**(3):431-437.

102.   Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing**. *Journal of the Royal Statistical Society, Series B (Methodological)* 1995, **57**(1):289-300.

103.   Curtis RK, Oresic M, Vidal-Puig A: **Pathways to the analysis of microarray data**. *Trends Biotechnol* 2005, **23**(8):429-435.

104.   Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nat Protoc* 2009, **4**(1):44-57.

105.   Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery**. *Genome Biol* 2003, **4**(5):P3.

106.   Revil T, Gaffney D, Dias C, Majewski J, Jerome-Majewska LA: **Alternative splicing is frequent during early embryonic development in mouse**. *BMC Genomics* 2010, **11**:399.

107.   Cline MS, Blume J, Cawley S, Clark TA, Hu JS, Lu G, Salomonis N, Wang H, Williams A: **ANOSVA: a statistical method for detecting splice variation from expression data**. *Bioinformatics* 2005, **21 Suppl 1**:i107-115.

108.   Purdom E, Simpson KM, Robinson MD, Conboy JG, Lapuk AV, Speed TP: **FIRMA: a method for detection of alternative splicing from exon array data**. *Bioinformatics* 2008, **24**(15):1707-1714.

109. Della Beffa C, Cordero F, Calogero RA: **Dissecting an alternative splicing analysis workflow for GeneChip Exon 1.0 ST Affymetrix arrays**. *BMC Genomics* 2008, **9**:571.

110. Clark TA, Schweitzer AC, Chen TX, Staples MK, Lu G, Wang H, Williams A, Blume JE: **Discovery of tissue-specific exons using comprehensive human exon microarrays**. *Genome Biol* 2007, **8**(4):R64.

111. Feingold KR, Zsigmond G, Lear SR, Moser AH: **Effect of food intake on intestinal cholesterol synthesis in rats**. *Am J Physiol* 1986, **251**(3 Pt 1):G362-369.

112. Walker BL, Hahn P: **Hepatic microsomal 3-hydroxy-3-methylglutaryl-CoA reductase activity in the lactating rat**. *Can J Biochem* 1981, **59**(11-12):889-892.

113. Belfiore A, Frasca F, Pandini G, Sciacca L, Vigneri R: **Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease**. *Endocr Rev* 2009, **30**(6):586-623.

114. Fernandez-Moreno MD, Arilla E, Prieto JC: **Insulin binding to rat intestinal epithelial cells following partial small-bowel resection**. *Biosci Rep* 1986, **6**(5):445-450.

115. Lostao MP, Urdaneta E, Martinez-Anso E, Barber A, Martinez JA: **Presence of leptin receptors in rat small intestine and leptin effect on sugar absorption**. *FEBS Lett* 1998, **423**(3):302-306.

116. Cohen P, Yang G, Yu X, Soukas AA, Wolfish CS, Friedman JM, Li C: **Induction of leptin receptor expression in the liver by leptin and food deprivation**. *J Biol Chem* 2005, **280**(11):10034-10039.

117. Lopez-Fontal R, Zeini M, Traves PG, Gomez-Ferreria M, Aranda A, Saez GT, Cerda C, Martin-Sanz P, Hortelano S, Bosca L: **Mice lacking thyroid hormone receptor Beta show enhanced apoptosis and delayed liver commitment for proliferation after partial hepatectomy**. *PLoS One*, **5**(1):e8710.

118. Kosters A, Karpen SJ: **Bile acid transporters in health and disease**. *Xenobiotica* 2008, **38**(7-8):1043-1071.

119. Wooton-Kee CR, Coy DJ, Athippozhy AT, Jones BR, Vore M: **Mechanisms for increased expression of cholesterol 7⟨-hydroxylase (Cyp7a1) in lactating rats**. *Hepatology* 2009.

120. Fernando RL, Nettleton D, Southey BR, Dekkers JC, Rothschild MF, Soller M: **Controlling the proportion of false positives in multiple dependent tests**. *Genetics* 2004, **166**(1):611-619.

121. Blalock EM, Chen KC, Sharrow K, Herman JP, Porter NM, Foster TC, Landfield PW: **Gene microarrays in hippocampal aging: statistical profiling identifies novel processes correlated with cognitive impairment**. *J Neurosci* 2003, **23**(9):3807-3819.

122. Arzuaga X, Ren N, Stromberg A, Black EP, Arsenescu V, Cassis LA, Majkova Z, Toborek M, Hennig B: **Induction of gene pattern changes associated with dysfunctional lipid metabolism induced by dietary fat and exposure to a persistent organic pollutant**. *Toxicol Lett* 2009, **189**(2):96-101.

123. Hulshizer R, Blalock EM: **Post hoc pattern matching: assigning significance to statistically defined expression patterns in single channel microarray data**. *BMC Bioinformatics* 2007, **8**:240.

124. Bartz F, Kern L, Erz D, Zhu M, Gilbert D, Meinhof T, Wirkner U, Erfle H, Muckenthaler M, Pepperkok R *et al*: **Identification of cholesterol-regulating genes by targeted RNAi screening**. *Cell Metab* 2009, **10**(1):63-75.

125. Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF *et al*: **Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences**. *Proc Natl Acad Sci U S A* 2002, **99**(26):16899-16903.

126. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T *et al*: **KEGG for linking genomes to life and the environment**. *Nucleic Acids Res* 2008, **36**(Database issue):D480-484.

127. Cunnane SC, Armstrong JK: **Long-chain fatty acid composition of maternal liver lipids during pregnancy and lactation in the rat: comparison of triglyceride to phospholipid**. *J Nutr* 1990, **120**(4):338-345.

128. Russell DW: **The enzymes, regulation, and genetics of bile acid synthesis**. *Annu Rev Biochem* 2003, **72**:137-174.

129. Davis HR, Jr., Zhu LJ, Hoos LM, Tetzloff G, Maguire M, Liu J, Yao X, Iyer SP, Lam MH, Lund EG *et al*: **Niemann-Pick C1 Like 1 (NPC1L1) is the intestinal phytosterol and cholesterol transporter and a key modulator of whole-body cholesterol homeostasis**. *J Biol Chem* 2004, **279**(32):33586-33592.

130. Yu L, Li-Hawkins J, Hammer RE, Berge KE, Horton JD, Cohen JC, Hobbs HH: **Overexpression of ABCG5 and ABCG8 promotes biliary cholesterol secretion and reduces fractional absorption of dietary cholesterol**. *J Clin Invest* 2002, **110**(5):671-680.

131. Alvarez AI, Real R, Perez M, Mendoza G, Prieto JG, Merino G: **Modulation of the activity of ABC transporters (P-glycoprotein, MRP2, BCRP) by flavonoids and drug response**. *J Pharm Sci*, **99**(2):598-617.

132. Lichten LA, Cousins RJ: **Mammalian zinc transporters: nutritional and physiologic regulation**. *Annu Rev Nutr* 2009, **29**:153-176.

133. Tao L, Wadsworth S, Mercer J, Mueller C, Lynn K, Siekierka J, August A: **Opposing roles of serine/threonine kinases MEKK1 and LOK in regulating the CD28 responsive element in T-cells**. *Biochem J* 2002, **363**(Pt 1):175-182.

134. Shimano H, Horton JD, Shimomura I, Hammer RE, Brown MS, Goldstein JL: **Isoform 1c of sterol regulatory element binding protein is less active than isoform 1a in livers of transgenic mice and in cultured cells**. *J Clin Invest* 1997, **99**(5):846-854.

135. Espenshade PJ: **SREBPs: sterol-regulated transcription factors**. *J Cell Sci* 2006, **119**(Pt 6):973-976.

136. Chen G, Liang G, Ou J, Goldstein JL, Brown MS: **Central role for liver X receptor in insulin-mediated activation of Srebp-1c transcription and stimulation of fatty acid synthesis in liver**. *Proc Natl Acad Sci U S A* 2004, **101**(31):11245-11250.

137. Sakai J, Nohturfft A, Goldstein JL, Brown MS: **Cleavage of sterol regulatory element-binding proteins (SREBPs) at site-1 requires interaction with SREBP cleavage-activating protein. Evidence from in vivo competition studies**. *J Biol Chem* 1998, **273**(10):5785-5793.

138. Shimano H: **Sterol regulatory element-binding proteins (SREBPs): transcriptional regulators of lipid synthetic genes**. *Prog Lipid Res* 2001, **40**(6):439-452.

139. Chiang JY: **Bile acid regulation of gene expression: roles of nuclear hormone receptors**. *Endocr Rev* 2002, **23**(4):443-463.

140. Saher G, Brugger B, Lappe-Siefke C, Mobius W, Tozawa R, Wehr MC, Wieland F, Ishibashi S, Nave KA: **High cholesterol level is essential for myelin membrane growth**. *Nat Neurosci* 2005, **8**(4):468-475.

141. Clarenburg R, Chaikoff IL: **Origin of milk cholesterol in the rat: dietary versus endogenous sources**. *J Lipid Res* 1966, **7**(1):27-37.

142. Tang X, Shay NF: **Zinc has an insulin-like effect on glucose transport mediated by phosphoinositol-3-kinase and Akt in 3T3-L1 fibroblasts and adipocytes**. *J Nutr* 2001, **131**(5):1414-1420.

143. Rink L, Kirchner H: **Zinc-altered immune function and cytokine production**. *J Nutr* 2000, **130**(5S Suppl):1407S-1411S.

144. Krebs NF: **Zinc supplementation during lactation**. *Am J Clin Nutr* 1998, **68**(2 Suppl):509S-512S.

145. James K: **Interactions between cytokines and alpha 2-macroglobulin**. *Immunol Today* 1990, **11**(5):163-166.

146. Borth W, Luger TA: **Identification of alpha 2-macroglobulin as a cytokine binding plasma protein. Binding of interleukin-1 beta to "F" alpha 2-macroglobulin**. *J Biol Chem* 1989, **264**(10):5818-5825.

147. Ho RH, Kim RB: **Transporters and drug therapy: implications for drug disposition and disease**. *Clin Pharmacol Ther* 2005, **78**(3):260-277.

148. Andric SA, Kostic TS, Stojilkovic SS: **Contribution of multidrug resistance protein MRP5 in control of cyclic guanosine 5'-monophosphate intracellular signaling in anterior pituitary cells**. *Endocrinology* 2006, **147**(7):3435-3445.

149. Beck K, Hayashi K, Dang K, Hayashi M, Boyd CD: **Analysis of ABCC6 (MRP6) in normal human tissues**. *Histochem Cell Biol* 2005, **123**(4-5):517-528.

150. Belinsky MG, Chen ZS, Shchaveleva I, Zeng H, Kruh GD: **Characterization of the drug resistance and transport properties of multidrug resistance protein 6 (MRP6, ABCC6)**. *Cancer Res* 2002, **62**(21):6172-6177.

151. van Herwaarden AE, Schinkel AH: **The function of breast cancer resistance protein in epithelial barriers, stem cells and milk secretion of drugs and xenotoxins**. *Trends Pharmacol Sci* 2006, **27**(1):10-16.

152. Fukuda H, Ohshima K, Mori M, Kobayashi I, Greer MA: **Sequential changes in the pituitary-thyroid axis during pregnancy and lactation in the rat**. *Endocrinology* 1980, **107**(6):1711-1716.

153. Li H, Wood CL, Getchell TV, Getchell ML, Stromberg AJ: **Analysis of oligonucleotide array experiments with repeated measures using mixed models**. *BMC Bioinformatics* 2004, **5**:209.

154. **The Universal Protein Resource (UniProt) in 2010**. *Nucleic Acids Res*, **38**(Database issue):D142-148.

155. Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, Suzek BE, Martin MJ, McGarvey P, Gasteiger E: **Infrastructure for the life sciences: design and implementation of the UniProt website**. *BMC Bioinformatics* 2009, **10**:136.

156. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L *et al*: **InterPro: the integrative protein signature database**. *Nucleic Acids Res* 2009, **37**(Database issue):D211-215.

157. Wu CH, Huang H, Yeh LS, Barker WC: **Protein family classification and functional annotation**. *Comput Biol Chem* 2003, **27**(1):37-47.

158. Schultz J, Milpetz F, Bork P, Ponting CP: **SMART, a simple modular architecture research tool: identification of signaling domains**. *Proc Natl Acad Sci U S A* 1998, **95**(11):5857-5864.

159. Letunic I, Doerks T, Bork P: **SMART 6: recent updates and new developments**. *Nucleic Acids Res* 2009, **37**(Database issue):D229-232.

160. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res* 2000, **28**(1):27-30.

161. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG**. *Nucleic Acids Res* 2006, **34**(Database issue):D354-357.

162. Zhang Y, Zhao T, Li W, Vore M: **The 5'-untranslated region of multidrug resistance associated protein 2 (MRP2; ABCC2) regulates downstream open reading frame expression through translational regulation**. *Mol Pharmacol*, **77**(2):237-246.

163. Andersen CL, Jensen JL, Orntoft TF: **Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets**. *Cancer Res* 2004, **64**(15):5245-5250.

164. Athippozhy A, Huang L, Wooton-Kee CR, Zhao T, Jungsuwadee P, Stromberg AJ, Vore M: **Differential gene expression in liver and small intestine from lactating rats compared to age-matched virgin controls detects increased mRNA of cholesterol biosynthetic genes**. *BMC Genomics* 2011, **12**(1):95.

165. Affymetrix I: **Alternative Transcript Analysis Methods for Exon Arrays**. In.; 2005.

166. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC**. *Genome Res* 2002, **12**(6):996-1006.

167. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A *et al*: **The UCSC Genome Browser database: update 2011**. *Nucleic Acids Res* 2011, **39**(Database issue):D876-882.

168. Hirata T, Okabe M, Kobayashi A, Ueda K, Matsuo M: **Molecular mechanisms of subcellular localization of ABCG5 and ABCG8**. *Biosci Biotechnol Biochem* 2009, **73**(3):619-626.

169. Kiang TK, Ensom MH, Chang TK: **UDP-glucuronosyltransferases and clinical drug-drug interactions**. *Pharmacol Ther* 2005, **106**(1):97-132.

170. Chen S, Beaton D, Nguyen N, Senekeo-Effenberger K, Brace-Sinnokrak E, Argikar U, Remmel RP, Trottier J, Barbier O, Ritter JK *et al*: **Tissue-specific, inducible, and hormonal control of the human UDP-glucuronosyltransferase-1 (UGT1) locus**. *J Biol Chem* 2005, **280**(45):37547-37557.

171. Hamming I, Cooper ME, Haagmans BL, Hooper NM, Korstanje R, Osterhaus AD, Timens W, Turner AJ, Navis G, van Goor H: **The emerging role of ACE2 in physiology and disease**. *J Pathol* 2007, **212**(1):1-11.

172. Carey RM: **Aldosterone and cardiovascular disease**. *Curr Opin Endocrinol Diabetes Obes* 2010, **17**(3):194-198.

173. Gallagher PE, Chappell MC, Ferrario CM, Tallant EA: **Distinct roles for ANG II and ANG-(1-7) in the regulation of angiotensin-converting enzyme 2 in rat astrocytes**. *Am J Physiol Cell Physiol* 2006, **290**(2):C420-426.

174. Kent WJ: **BLAT--the BLAST-like alignment tool**. *Genome Res* 2002, **12**(4):656-664.

175. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE *et al*: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution**. *Nature* 2004, **428**(6982):493-521.

176. Tukey RH, Strassburg CP: **Human UDP-glucuronosyltransferases: metabolism, expression, and disease**. *Annu Rev Pharmacol Toxicol* 2000, **40**:581-616.

177. Vrzal R, Ulrichova J, Dvorak Z: **Aromatic hydrocarbon receptor status in the metabolism of xenobiotics under normal and pathophysiological conditions**. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2004, **148**(1):3-10.

178. Tijet N, Boutros PC, Moffat ID, Okey AB, Tuomisto J, Pohjanvirta R: **Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries**. *Mol Pharmacol* 2006, **69**(1):140-153.

179. Tekpli X, Rissel M, Huc L, Catheline D, Sergent O, Rioux V, Legrand P, Holme JA, Dimanche-Boitrel MT, Lagadic-Gossmann D: **Membrane remodeling, an early event in benzo[a]pyrene-induced apoptosis**. *Toxicol Appl Pharmacol* 2010, **243**(1):68-76.

180. Hiller M, Platzer M: **Widespread and subtle: alternative splicing at short-distance tandem sites**. *Trends Genet* 2008, **24**(5):246-255.

181. Partek_Inc.: **Partek Genomics Suite Version 6.3**. In. St. Louis, MO: Partek Inc.; 2008.

182. Rishi V, Gal J, Krylov D, Fridriksson J, Boysen MS, Mandrup S, Vinson C: **SREBP-1 dimerization specificity maps to both the helix-loop-helix and leucine zipper domains: use of a dominant negative**. *J Biol Chem* 2004, **279**(12):11863-11874.

183. Zoumi A, Datta S, Liaw LH, Wu CJ, Manthripragada G, Osborne TF, Lamorte VJ: **Spatial distribution and function of sterol regulatory element-binding protein 1a and 2 homo- and heterodimers by in vivo two-photon imaging and spectroscopy fluorescence resonance energy transfer**. *Mol Cell Biol* 2005, **25**(8):2946-2956.

184.   Anderson SM, Rudolph MC, McManaman JL, Neville MC: **Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis!** *Breast Cancer Res* 2007, **9**(1):204.

185.   Shimomura I, Bashmakov Y, Ikemoto S, Horton JD, Brown MS, Goldstein JL: **Insulin selectively increases SREBP-1c mRNA in the livers of rats with streptozotocin-induced diabetes**. *Proc Natl Acad Sci U S A* 1999, **96**(24):13656-13661.

186.   Azzout-Marniche D, Becard D, Guichard C, Foretz M, Ferre P, Foufelle F: **Insulin effects on sterol regulatory-element-binding protein-1c (SREBP-1c) transcriptional activity in rat hepatocytes**. *Biochem J* 2000, **350 Pt 2**:389-393.

187.   Kambe T, Andrews GK: **Novel proteolytic processing of the ectodomain of the zinc transporter ZIP4 (SLC39A4) during zinc deficiency is inhibited by acrodermatitis enteropathica mutations**. *Mol Cell Biol* 2009, **29**(1):129-139.

188.   Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, Veitch J, Schweitzer A, Awad T, Sugnet C, Dee S *et al*: **Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array**. *BMC Genomics* 2006, **7**:325.

189.   Laajala E, Aittokallio T, Lahesmaa R, Elo LL: **Probe-level estimation improves the detection of differential splicing in Affymetrix exon array studies**. *Genome Biol* 2009, **10**(7):R77.

190.   Ron M, Israeli G, Seroussi E, Weller JI, Gregg JP, Shani M, Medrano JF: **Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle**. *BMC Genomics* 2007, **8**:183.

Antony Thomas Athippozhy

Born: March 24, 1983 Evansville, IN

Objective: Postdoctoral training in genomics and high throughput data anlaysis.

Education

August 2005-Present: PhD candidate, Graduate Center for Toxicology, University of Kentucky. Degree expected Summer, 2011.

August 2001-May 2005- B.S (Chemistry) University of Southern Indiana

Research Publications:

Antony Athippozhy, Liping Huang, Clavia Ruth Wooton-Kee, Tianyong Zhao, Paiboon Jungsuwadee, Arnold J Stromberg, Mary Vore *Differential gene expression in liver and small intestine from lactating rats compared to age-matched virgin controls detects increased mRNA of cholesterol biosynthetic genes. BMC Genomics* 2011, **12**:95 (3 February 2011)

Wooton-Kee CR, Coy DJ, **Athippozhy AT**, Zhao T, Jones BR, Vore M.

Mechanisms for increased expression of cholesterol 7alpha-hydroxylase

(Cyp7a1) in lactating rats. Hepatology. 2010 Jan;51(1):277-85.