PROTEOMIC ANALYSIS OF MAMMARY EPITHELIAL CELL DEVELOPMENT

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Agriculture, with a Specialization in Animal Science

by

Alyssa Kay Conly

January 2014

COMMITTEE MEMBERSHIP

TITLE:	Proteomic Analysis of Mammary Epithelial Cell Development
AUTHOR:	Alyssa Kay Conly
DATE SUBMITTED:	January 2014
COMMITTEE CHAIR:	Daniel G. Peterson, PhD Professor of Animal Science
COMMITTEE MEMBER:	Russel C. Hovey, PhD Professor of Animal Science, UC Davis
COMMITTEE MEMBER:	Lars Tomanek, PhD Associate Professor of Biological Sciences
COMMITTEE MEMBER:	Alexander Dekhtyar, PhD Associate Professor of Computer Science

ABSTRACT

Proteomic Analysis of Mammary Epithelial Cell Development

Alyssa Kay Conly

In this set of studies, a proteomic approach was used to investigate the protein profile of the mammary epithelial cell (MEC) through different stages of mammary development. The HC11 cell line was used to investigate protein changes between undifferentiated and differentiated MEC, which represent the pregnant and lactating states of the cells. This comparison revealed an interesting differential expression profile underscoring many recognized processes that occur in differentiated MECs, while others unveiled differences between MEC differentiation *in vitro* and *in vivo*. Primary MEC were also isolated from virgin, pregnant, and primiparous quiescent mice to compare the virgin state of the cell to the other two stages of development. These comparisons added to a previous dataset of primary isolated MEC and generated data that implied a surprising level of activity in virgin MEC relative to the other stages of development investigated. Differentially expressed proteins in the virgin and primiparous quiescent comparison also added to evidence of persisting changes occurring in the gland after a full term pregnancy that are implicated in the risk for breast cancer development. Data sets generated in the same manner from differentiating MEC were used in the development of a database to help manage the growing list of differentially expressed proteins and aid in the identification of potential interesting patterns of regulation during mammary development and differentiation.

ACKNOWLEDGMENTS

I would like to thank all of my committee members for being so accommodating and supportive. Also, I would like to thank my lab-mates Laura Strand and Nicole Einfalt for the help provided over the years and for keeping time in the lab entertaining. Marcus Zuzow and the students in the Tomanek lab also deserve many thanks for all of the help they have provided over the past couple of years. Lastly, I would like to especially thank Dan Peterson for all of his guidance and for his contribution to making my experience at Cal Poly a wonderful one.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 – Literature Review

1.1 Introduction

At the center of milk production lays the secretory mammary epithelial cell (MEC). This cell is responsible for the production and secretion of all milk components that are relied upon to solely nourish newborn offspring of the Class Mammalia. With changing stages of the reproductive life of the female mammal, the role of the mammary gland also changes. As the gland prepares for lactation the MEC population must expand and develop into cells capable of the metabolic power necessary to fulfill their function. The hormonal cues of puberty, pregnancy and lactation influence the development and differentiation of the MEC into a milk producing phenotype. Though the changing hormonal cues playing a role in the development of the mammary gland are well documented, the resulting molecular signals are less understood. This topic is being heavily investigated as better understanding of these underlying mechanisms can lead to improvements in the dairy industry, breast feeding, and even prevention and treatment of breast cancer.

Approaches to the study of biological science fields, such as mammary development, have shifted to the evaluation of whole systems and networks of data at once by performing experiments that allow discovery of new molecules and processes instead of confirmation or refutation of theories about a specific pathway. The broad topics of study focused on whole systems have become to be known as the "-omics." These studies allow the discovery of new investigative topics and the ability to see more factors interacting in a system at one time. These methods require collaboration with technology-related fields in order to produce computer systems able to manage and analyze the large and complex

datasets being generated. An interdisciplinary approach is recognized as essential and is being implemented as shown by the emerging field of bioinformatics.

This review will cover the stages of development of the mammary gland, the mechanisms behind its changing phenotype that have been discovered thus far, and importance of the emergence of interdisciplinary studies in advancing understanding of systems such as this one.

1.2 The Mammary Gland

1.2.1 Anatomy of the fully developed gland

A unique characteristic of the mammary gland is its ability to repeatedly undergo functional differentiation and involution to quiescence throughout the sexually productive period in the life of the female mammal. The majority of the gland's development occurs beginning at puberty, but it is not fully developed and functional until the female experiences pregnancy and lactation. The gland is comprised of two tissue compartments: the epithelial parenchyma and the stroma. This section describes the anatomy and basic function of these compartments as they exist in a fully developed mammary gland in order to provide understanding of all structures and cells that can be found in the gland. The following section will walk through specifics for each stage of mammary gland development.

1.2.1.1 Parenchyma

The parenchyma is the tissue of the gland that is responsible for the synthesis, secretion and transport of milk to the offspring. Its overall structure resembles a tree branching into the bed of the stromal tissue of the gland. This branching structure varies slightly among species, but its general components and functions remain the same.

1.2.1.1.1 The alveolus

When following a drop of milk through the mammary gland from its origin of production to its expulsion into the mouth of the offspring, the journey begins in a structure called the alveolus, or acinus. Alveoli are small spherical structures made up of a lumen surrounded by two different layers of distinct cell types and are often clustered in bunches called lobules. Looking from the inside of the lumen, secretory luminal epithelial cells are seen lining the inner layer of the alveolus. The secretory MEC are cuboidal to columnar in shape and are connected to each other by tight junctions, allowing strict control over everything that enters or leaves the lumen of the alveolus. These tight junctions change in permeability based on different stimuli, but are impermeable during lactation to prevent paracellular leaking in either direction in order to maintain milk composition and to allow storage between times of nursing (Nguyen and Neville, 1998; Mepham, 1983). These are the functional cells of milk production; secretory MEC are responsible for all synthesis and transport of the necessary components into the alveolar lumen to create the final product known as milk. The processes underlying synthesis of these different milk components by the secretory MEC will be discussed later in this chapter.

Unseen from the inside, but sprawled around the outside of this layer of secretory cells in a pattern resembling a woven basket is the branching basal myoepithelial cell layer. Lactating rat mammary glands were found to have 4-6 myoepithelial cells per alveolus with each cell having 8-10 terminally branched processes, allowing encompassment of the structure (Nagato et al., 1980). These cells resemble smooth muscle cells in their filament profile and have the ability to contract in response to the hormone oxytocin, causing them to compress the alveolus and force milk out of the lumen into the connecting duct (Emerman and Vogl, 1986; Nagato et al., 1980). Myoepithelial cells contain gap junctions between each other, most likely for coordination of this contraction (Mepham, 1983).

Though its size may vary, the alveolus is present and similarly structured in mice, cattle and humans and performs the same essential functions.

1.2.1.1.2 The ductal system

When the signal is given by oxytocin, the alveolar myoepithelial cells contract and the milk is pushed from the alveolar lumen into the ductal system. The ductal system transports the milk unaltered from the site of production to the site of ejection. Smaller ducts collect milk from the alveolar lumen and transport it to progressively larger ducts, eventually culminating in a single collecting duct that terminates at an orifice in the nipple or teat (Mepham, 1983). Arrangement, number and size of the ducts vary with species, but the microstructure is conserved.

The ducts transporting the milk through the gland are very similar to the alveolus in cellular composition. Cuboidal to columnar epithelial cells resembling the secretory MEC line the ductal lumen, but these epithelial cells do not produce and secrete milk components like those of the alveoli (Hassiotou and Geddes, 2013). Tight junctions are still present and very important in the ductal lumen in order to maintain the pressure needed to transport and eject the milk from the gland as well as prevent any movement of material in and out of the ductal system through the paracellular route (Nguyen and Neville, 1998).

A myoepithelial layer also surrounds this layer of non-secretory luminal epithelial cells. Unlike the branching basket-like pattern of the alveolar myoepithelium, the cells of the duct are long and narrow in shape, arranged parallel to the direction of milk flow through the duct. Contraction of these cells results in a decrease in length and increase in diameter of the duct, allowing the quick movement of a larger volume of milk through the gland and to the nipple for ejection in response to the oxytocin cue (Emerman and Vogl, 1986).

Ductal microstructure and function is conserved greatly across species, but patterns of branching and numbers of ducts vary among them. Female mice have only one ductal opening at each nipple. Each duct divides into secondary ducts and forms a branching tree network ending in clusters of alveoli, but all milk produced in each gland culminates into the single lactiferous duct leading to each nipple (Cardiff and Allison, 2012). Like the female mouse, the cow has a branching epithelium draining into only one lactiferous duct leading to each teat. While the mouse and the human have ducts with relatively constant diameter leading up to the point of milk ejection and little room in the way of milk storage, the cow has an enlargement in the diameter of the duct where the milk from

all the branches collects prior to the duct orifice. This enlarged portion of the duct is used for milk storage and is called the gland cistern (Reece, 2009). Unlike the cow and mouse, the human has multiple ducts associated with a single mammary gland reaching the surface of each nipple. Each duct leads to a completely separated branching lobe of the mammary gland divided by fibrous stroma. The number of ducts and lobes per human breast varies greatly among individuals, where the most recent data reports an average of nine fully functional lobes with ducts opening to the surface (Ramsay et al., 2005; Gooding et al., 2010; Hassiotou and Geddes, 2013). While they vary in organization, the ducts for all three have their individual benefits and satisfy the purpose of the ductal system, which is transport of the final unaltered milk product from the site of production to the site of ejection.

1.2.1.1.3 The teat

After its journey through the ducts, the milk arrives at the nipple or teat where the offspring can latch on and consume the ejected milk from the gland. Though not truly considered part of the parenchyma, the nipple plays a large role in the final step of milk delivery to the young and warrants discussion alongside the parenchymal components. This region is characterized by an outside covering of hairless epidermis formed around the orifice of the lactiferous duct, or ducts (Koyama et al., 2013). Smooth muscle is also found in the region near the epidermis and near the ducts and is important for the protrusion of the nipple to allow latching of offspring and for sphincter like function to prevent milk escaping from the teat between times of nursing (Hassiotou and Geddes, 2013).

Not only is this the site of milk expulsion and structurally important for the delivery of milk to offspring, but it is the site of afferent nerve endings that influence every event in the process of nursing young. Stimulation of these nerves causes release of oxytocin which cues the contraction of myoepithelium, as well as the release of prolactin, whose effects will be further covered later in this chapter (Koyama et al., 2013).

The nipples, or teats, of mammals vary in number and shape among species determined by the need of the offspring. Female mice have ten total nipples, each associated with its own mammary gland. They are positioned in pairs along the ventral portion of the mouse with three in the pectoral region and two in the inguinal region (Koyama et al., 2013). Cows have four total teats, also each associated with its own gland. They are positioned in the udder as two pairs, located in the inguinal region of the cow (Reece, 2009). Humans have only two nipples, each also associated with its own gland. As mentioned previously, each gland consists of multiple completely separated lobes that have individual openings at the same nipple, but they are still collectively considered a single gland. Position of the human glands exists as one pair in the pectoral region (Koyama et al., 2013).

1.2.1.2 Stroma

Parenchyma may produce and secrete and transport the milk components, but without the stroma the mammary gland would not be able to fully develop and carry out its intended function. As has been covered several reviews, stromal tissue of the mammary gland provides structural support for glandular tissue and delivers essential growth factors and hormones that direct the development and maintain the structure and function of the gland (McCave e al., 2010; Hovey et al., 1999). Components of the stroma include the extracellular matrix (ECM), adipocytes, immune cells, fibroblasts, and the endothelium. Each of these components varies in abundance and organization among species, but they all play an important role in the development and function of the mammary gland.

1.2.1.2.1 Extracellular matrix

The ECM is not a cellular component itself but is essential for and isproduced by the cells of the mammary gland. The ECM is composed of macromolecules secreted by the cells of the connective tissue and MEC. Glycosaminoglycan and fibrous proteins such as collagen, elastin, and laminin are the major components forming the matrix. The composition and organization of the matrix is constantly in flux and maintained by the cells of the stroma in response to released factors as well as mechanical stress (Alberts et al., 2002). Different regions of ECM are composed of different proportions of the mentioned proteins because of the role it needs to fill and the cell types present that are responsible for secreting the components (Maller et al., 2010). In humans and cows the ECM, along with fibroblasts, creates a thick fibrous layer separating the adipose tissue and parenchyma of the mammary gland and is extensive between and within the lobules

in these species to help provide extra support and maintain the form of these glands due to their size and placement (Hovey et al., 1999).

In addition to providing structural support to the parenchyma by allowing attachment and stability, the ECM has also been implicated in apoptosis, proliferation, branching, cell polarity and differentiation. The proteins of the ECM serve as signaling molecules in signal transduction pathways as well as mechanotransducers by connecting the ECM to the cytoskeleton of the cells (Butcher et al., 2009). Growth factors and cytokines are also bound to the proteins of the ECM, whereby it can serve as a reservoir for the sequestration of signaling molecules in specific locations for different purposes such as directing ductal growth during development or recruiting immune cells which are present and active in all stages of development (Maller et al., 2010).

1.2.1.2.2 Immune cells

Cells of the immune system are present and active throughout mammary development and function. Branching morphogenesis, terminal end bud (TEB) formation and ductal outgrowth are impaired without the presence of mast cells, macrophages and eosinophils. Lilla and Werb demonstrated decreased cell proliferation and impaired mammary development in mice deficient in or lacking mast cells (2010). Mast cells were seen concentrated near or ahead of the TEB in developing murine mammary glands. It was determined that mast cells most likely function in development by contribution of some component released during degranulation. Though uncertainty surrounding their exact function in development persists, it is speculated that released proteases or other factors aid in angiogenesis, remodeling the ECM or activating other proteases. Macrophages and

eosinophils are also known to exist in close proximity to the TEB during branching morphogenesis and ductal outgrowth and are necessary for normal development. Depletion of these cells resulted in the complete absence of TEBs and stunted ductal outgrowth (Gouon-Evans et al., 2002, 2000). Macrophage-specific depletion showed defective TEB formation and reduced ductal outgrowth indicating a role in these processes, probably accomplished by phagocytic activity and trophic factor release (Gouon-Evans et al., 2000). Eosinophil-specific depletion resulted in decreased number of branches indicating a role in branching of the ducts, possibly carried out by release of TGF-β or other factors (Gouon-Evans et al., 2000).

Development during pregnancy and lactation also requires the influence of immune cells. A need for colony stimulating factor 1 (CSF-1) has been described in the development of a properly functioning lobulo-alveolar system; in addition to a reduced number of macrophages, CSF-1 deficient mammary gland resulted in the inability to switch to the lactating phenotype, suggesting a possible role for macrophages in this process (Pollard and Henninghausen, 1994). In addition to the development, plasma cells are recruited to the mammary gland during lactation and are essential for production of antibodies that are incorporated into the milk by the epithelial cells. IgA is the primary antibody produced in the mammary gland and is found in the milk of mice, cattle and humans. Other antibodies are transferred into the milk from the blood including the main antibody present in bovine colostrum, IgG (Weisz-Carrington et al., 1977; Hunziker and Kraehenbul, 1998; Bourges et al., 2008). Production and delivery of these antibodies through the milk is essential to the first defense of the offspring.

In addition to the developmental functions mentioned above, the immune cells of the mammary gland are also fulfilling their most well known function of scouting the gland for infiltrators and keeping the tissue and the milk free from infection (Stelwagen et al., 2009).

1.2.1.2.3 Adipocytes

The mammary gland of the mouse contains a far higher proportion of adipocytes than the cow or human (Hovey et al., 1999). Many of the studies investigating the roles of these adipocytes have been performed with mice; applying conclusions from results across species should be implemented with caution. Though previously thought to play little role in the development and function of the mammary gland, essential functions of the mammary adipocytes have been unveiled, as nicely outlined in a recent review (Hovey and Aimo, 2010). Adipose tissue has been acknowledged as an indispensable component of the mammary stroma.

Of the many roles the adipocytes play in the gland, one is provision of nutrient components to the epithelial cells to aid in milk production during lactation. Depletion of the fat stores in adipocytes in the lactating mammary gland has been observed and is believed to provide fatty acids to the epithelium (Elias et al., 1973; Clegg, 1981), though it is believed that only a small proportion of the fatty acids needed by the MEC are supplied by the adipocytes (Neville et al., 1998). The rest would be synthesized in the MEC or delivered through the blood from diet or other fat deposits. Adipocytes themselves have also been found to release lipoprotein lipase into the circulation of the

mammary gland to breakdown TAG from the blood into fatty acids for the parenchyma to utilize to for milk component synthesis (Jensen et al., 1994).

In addition to providing nutrients, adipocytes have the ability to produce a multitude of paracrine and endocrine factors including growth factors and hormones. Among these molecules produced is the hormone prolactin, which was previously thought to have the sole production center in the anterior pituitary gland. This has been found true in humans and mice, though its autocrine or paracrine role has not been confirmed as necessary (Zinger et al., 2003). Potential for estrogen production in the mammary adipose due to expression of aromatase has also been found, which was shown to be prolactin-inducible in bovine mammary adipose cultures (Feuerrman et al., 2009). Many growth factors are also produced by the adipocytes of the mammary gland such as insulin-like growth factor 1 (IGF-1), transforming growth factor-β (TGF-β), and vascular endothelial growth factor (VEGF), which take part in development of the glandular epithelium, response to hormone activation, and development of the mammary vasculature, respectively.

1.2.1.2.4 Fibroblasts

Fibroblasts are the major cells of connective tissue and secrete the components making up much of the ECM, including production of components in the basement membrane such as laminin, nidogens, and collagen. The ECM, as mentioned above, is essential for support of the glandular tissue and proper development and function of the mammary gland; presence of some of these components has been shown necessary for polarization and differentiation of mammary epithelium (Luhr et al., 2012). Fibroblasts are also

responsible for the secretion of growth factors and matrix metalloproteinases that play roles in multiple stages of mammary development (Simian et al., 2001).

1.2.1.2.5 Endothelium

Vascularization of the mammary gland is present throughout the development of the organ, with vast expansion occurring during pregnancy and regression at involution (Andres and Djonov, 2010). Endothelial cell precursors are recruited from the bone marrow or created from resident pluripotent stem cells in the tissue to form and remodel microvasculature during pregnancy by the influence of hormones such as estrogen (Kabmeyer et al., 2009; Cid et al., 2002). During pregnancy and lactation, the endothelium is organized in capillaries enveloping each individual alveolus and the capillaries become more permeable to enable efficient nutrient transfer (Djonov et al., 2001; Abdul-Awal et al., 1996). In humans it was found that relatively large and few sinusoidal shaped capillaries encircle each alveolus to allow slow movement of blood and ensure transport of essential nutrients and hormonal signals to the alveolar cells (Naccarato et al., 2003). The major known molecule playing a large role in vasculature forming and remodeling during pregnancy is VEGF, which has been shown to be inducible by estrogen and is secreted by the epithelium and other cells of the stroma (Rabbany et al., 2003; Hovey et al., 2001; Andres and Djonov, 2010).

1.2.2 Developmental events and stages of the mammary gland.

1.2.2.1 Pubertal growth

Until puberty the mammary gland undergoes isometric growth, at which time females experience allometric growth of the gland in response to changing hormone levels. A rise in gonadotropin releasing hormones from the newly mature hypothalamic-pituitary-gonadal axis at the onset of puberty causes production of ovarian hormones (Howlin et al., 2006; Plant and Barker-Gibbs, 2004). The production and release of these hormones cause the outgrowth of the mammary ductal tree into the surrounding stroma.

In the rodent, highly proliferative bulb-shaped structures called terminal end buds (TEB) are formed in the parenchyma and lead the ducts as they extend through the existing fat pad. The TEB is made up of different layers of epithelial cells with the outermost layer being made up of a stem cell population referred to as the cap cells. These cells line the outside of the TEB from the leading edge into the neck region of the structure and are continuous with the differentiated myoepithelial layer of the duct, as these cap cells eventually give rise to the myoepithelium (Richert et al., 2000; Williams and Daniel, 1983). At the trailing edge of the TEB, the differentiated myoepithelial cells produce basement membrane surrounding the duct while the fibroblasts of the fat pad secrete other components of the stroma surrounding the membrane (Daniel and Silberstein, 1987; Williams and Daniel, 1983). This outside layer of cap cells surrounds several layers of epithelial cells referred to as the body cells. These cells eventually become the ductal epithelial cells (Daniel and Silberstein, 1987). The innermost layer of body cells undergoes apoptosis in order to form the lumen of the duct; in fact, the levels of apoptosis observed in the TEB during ductal expansion (11.3%) are higher than observed levels in

any other stage of mammary development, including involution (4%; Quarrie et al., 1995; Humphreys et al., 1996). Bifurcation of the TEB allows for full outspread to the edges of the fat pad. Though it is difficult to study pubertal development of the mammary gland in humans, it appears that there are also end bud-like structures leading the expansion of the ducts during this period of outgrowth, though cap cells like in rodent TEB have not been identified (Howard and Gusterson, 2000). Bovine mammary glands are a bit different in their structure and timing of outgrowth with a central, solid epithelial cord surrounded by 5-10 ductal outgrowths called a terminal ductal unit (TDU) that influence allometric growth prior to puberty. The whole unit invades the stroma together and the lumen does not appear to be created by high levels of apoptosis as has been determined in the rodent gland (Capuco and Ellis, 2005). Hormone levels, the hormone receptor profile of different cell types, and growth factors within the mammary gland dictate these patterns of growth.

Among several hormones affecting this surge in growth, estrogen (E) has the primary effect. Estrogen receptor α (ERα) is essential for ductal outgrowth during pubertal development, while ERβ doesn't seem to play a role (Mallepell et al., 2006). Presence of ERα in the mammary epithelium was shown to be necessary for this development while its presence in the stroma did not affect the ductal outgrowth (Mallepell et al., 2006). Cells devoid of ERα were shown to proliferate - in fact, in rodents and humans proliferating cells rarely contain steroid receptors (Clarke et al., 1997; Russo et al., 1999). It is thought that E exerts its effects in a paracrine manner, acting only on a subset of cells in the mammary epithelium which then produce and secrete additional factors that act on other epithelial cells directly, or on stromal cells, also causing release of growth factors

that would influence epithelial outgrowth. Amphiregulin (AREG) is thought to be the molecule responsible for mediating this response to E. Its receptor, epidermal growth factor receptor (EGFR), is necessary in the fat pad for normal development and stromal cells secrete growth factors in response to this molecule (Wiesen et al., 1999; Sternlicht et al., 2005). Though the increase of ovarian E at puberty is of utmost importance, it is well documented that the presence of a certain pituitary hormone is essential for ductal outgrowth as well (Sternlicht, 2005).

Growth hormone (GH) is also necessary for ductal outgrowth during puberty. It is thought to influence the formation of the TEB that begins invading the mammary stroma (Kleinberg, 1997). Growth hormone is produced by the anterior pituitary and acts on the stroma of the mammary gland to induce expression of ER and insulin-like growth factor 1 (IGF-1). This locally produced factor then acts on the MEC to induce formation of TEB, and proliferation and outgrowth of the ductal system, which without the ability to respond to IGF-1 does not occur (Walden et al., 1998; Wood et al., 2000; Hadsell and Bonnette, 2000). Estrogen acts synergistically with GH and IGF-1 to allow proper outgrowth and development of the pubertal gland (Kleinberg, 1997; Ruan et al., 1995).

Far less credit has been given to other hormones and steroid receptors, and while E and GH remain to be considered the hormones most involved in ductal outgrowth during puberty, a few others have been implicated in the regulation of this process as well and should be given an honorable mention. Though little evidence has been found for a large role for progesterone (P) in the pubertal mammary gland, it has been shown that pubertal branching is delayed when PR is antagonized (Shi et al., 2004) and it was recently shown that progesterone might also work through AREG to aid in ductal growth and branching

in the pubertal gland (Auperlee et al., 2013). Though these new functions are being uncovered, the largest role P plays still occurs during pregnancy. Knockout of the glucocorticoid receptor has also caused abnormal development during puberty, though this is difficult to properly study because of the necessity of transplanting mammary tissue from the embryo into fat pads due to lethal consequences of the deletion (Kingsley-Kallesen et al., 2002). In contrast to these previous two, the vitamin D3 receptor (VDR) seems to aid in the repression of ductal outgrowth. Mice with VDR knockouts displayed accelerated outgrowth and ductal branching during puberty (Zinser et al., 2002). It is speculated that E causes transcription and translation of VDR, which appears to antagonize the proliferation of the epithelium, possibly through the regulation of the ER itself (Byrne et al., 2000; Welsh et al., 2003; Howlin et al., 2006).

In addition to hormones and steroid receptors, additional locally produced growth factors much like IGF-1 also play a role in directing and regulating mammary ductal outgrowth. Due to the focus of this review, these factors will not be discussed but their important roles in development are continuously being discovered. These include factors such as epidermal growth factor, TGF-β, and hepatic growth factor. This list is not exhaustive and full understanding on their effects of the gland and its development has not been achieved. While it is clear that E and GH drive the outgrowth and development of the epithelium, the regulation of proper growth and development during puberty is optimal when many different hormones, receptors and factors are present and working together in harmony.

1.2.2.2 Virgin adult

Terminal end buds of the developing ducts of the mammary gland have reached the borders of the fat pad and regressed to non-motile units by the end of puberty. Although full development and capacity to secrete milk components does not occur until pregnancy, changes in the gland occur with every estrous cycle. Tertiary side branches begin to expand within the stroma between developed ducts and acini and show development and regression with the cyclic production of different ovarian hormones (Daniel and Silberstein, 1987). In both rodents and humans the mammary epithelium cycles through stages of proliferation, differentiation and apoptosis that mirror the hormone changes through the estrous cycle. Peak proliferation of MEC occurs during metestrus and diestrus-1 in rodents and the equivalent, late luteal phase, in humans (Schedin et al., 2000). Peak differentiation has been seen during the same time period of luteal phase and metestrus/diestrus-1, as measured by beta-casein and whey acid protein mRNA levels throughout the cycle (Schedin et al., 2000). Lastly, apoptosis levels were elevated and suppressed at similar points in the cycles of rodents and humans. Apoptosis was found to be much more variable than proliferation but was seen to be suppressed during times of peak proliferation and increased during proestrus and diestrus-2 (Schedin et al., 2000). Although they are continuously developing, the alveolar development never reaches near the level of differentiation of pregnancy.

1.2.2.3 Pregnancy

Though developing and regressing with each estrous cycle during the gland's nulliparous adult existence, the gland does not begin down the path of full development until conception. Prior to conception, the gland has branched via TEBs to the ends of the mammary fat pad and lacks only the extensive side branches and alveoli that are seen in a lactating mammal. After conception, the ovum implants in the uterus and causes hormonal changes that alert the body of pregnancy. In the human, cells that are destined to be the placenta begin producing human chorionic gonadotropin (hCG), which is responsible for the maintenance of the corpus luteum and its continued P and E secretions during early pregnancy. In rodents, prolactin (PRL) secretions are responsible for this early corpus luteum maintenance (Neville et al., 2002). The lengthened survival of the corpus luteum allows for its further production of P and E, which cause side branching and alveologenesis in the mammary gland and production of PRL from the anterior pituitary, respectively. At mid pregnancy placental lactogen takes over this role of corpus luteum maintenance in rodents while in humans, the placenta itself begins to produce the required hormones for maintenance of pregnancy and mammary development (Neville et al., 2002). While E induces expression of PR and PRLR in the mammary epithelium as well as production of PRL from the anterior pituitary, P and PRL work together to prepare for lactation. Both P and PRL each induce the expression of the other's receptor and individually work in a paracrine fashion, recruiting other factors to signal for ductal side branching, proliferation and MEC differentiation (Brisken et al., 1999; Edery et al., 1985; Sakai et al., 1979).

Progesterone acts through progesterone receptor B (PRB) to elicit responses related to

mammary development; it has been found responsible for side branching and mitotic activity of the epithelium during pregnancy (Brisken et al., 1998). As the majority of proliferating cells are PR negative, the possible paracrine activation of these cells has received much attention. Investigations into this topic are far from over, but candidate pathways have been proposed. Wnt-4 was the first proposed mediator of P action in the mammary gland. Brisken and colleagues showed that P induces Wnt-4 expression during early pregnancy and that it is capable of causing side branching in the mammary epithelium (2000) and Ramamoorthy and colleagues showed that Wnt-4 is a direct transcriptional target of PR (2010). Whether it causes a response in neighboring cells remains to be determined. Receptor activator of NF-κB-ligand (RANKL) and its receptor, RANK, are the strongest current candidates for the paracrine mediators of P action. Investigation into this factor began when Fata and colleagues discovered that mice lacking RANKL or its receptor failed to develop alveolar structures during pregnancy (2000). Since this discovery, RANKL has been found to be stimulated by P as well as have the ability to induce Elf5 expression (Beleut et al., 2010). This was particularly interesting because Elf5 has been found necessary for successful alveolar morphogenesis and lactation (Oakes et al., 2008; Choi et al., 2009). Elf5 and PR expression were found to be mutually exclusive so Lee and colleagues proposed that RANKL acts as a paracrine mediator for P in order to induce Elf5 (2013). RANKL is also thought to mediate a proliferation response to P in neighboring PR negative cells by inducing cyclin D1 expression (Fernandez-Valdivia et al., 2009). As this information is very recent, there is still much to learn about the intracellular signaling resulting in developmental changes in response to PR activation, but light has been shed on the role P has on development

during pregnancy and possible mediators of this response. In addition to P, PRL has an equally important role during pregnancy.

Prolactin and its receptor are necessary for normal alveolar development, full differentiation of the mammary epithelium, and successful lactation following parturition (Ormandy et al., 1997). Prolactin binds the prolactin receptor (PRLR), which activates the Janus kinase/signal transducer and activator of transcription (JAK/STAT) and protein kinase B (PKB/Akt) pathways. Through these pathways PRL affects structural development of the mammary gland during early pregnancy and production of milk components during late pregnancy and lactation. Signal transducer and activator of transcription 5 (STAT5) is a transcription factor in the PRLR/JAK/STAT pathway whose phosphorylated state can activate transcription of several genes involved in processes important for epithelial cell differentiation such as polarity establishment, epithelial cell-cell and epithelial-stromal interactions, and milk protein production (Oakes et al., 2006). Alveolar defects are seen in $STAT5^{-/-}$ mice and complete absence of alveolar structures and beta-casein is seen in $PRLR^{-/-}$ mice after full term pregnancy (Liu et al., 1997; Brisken et al., 1999). Much overlap has been seen in the intracellular signaling cascades that result from PR and PRLR activation. As stated above, PR can induce expression of Elf5 through RANKL. Elf5 seems to have the ability to induce expression of STAT5 and inhibit members of the suppressor of cytokine signaling (SOCS) family, which are inhibitors in the JAK/STAT phosphorylation pathway. Induction of Elf5 expression would both cause production of STAT5 and allow increased phosphorylation of this molecule by JAK in the PRLR/JAK/STAT pathway (Choi et al., 2009). There are conflicting studies that report whether or not PRLR might also induce RANKL through

the JAK/STAT pathway and play a role in activating Akt pathways that decrease apoptosis and delay involution (Srivastava et al., 2003; Brisken et al., 2002). Another molecule speculated to play a role in PRLR signaling is IGF-2. Activated PRLR induces expression of IGF-2, which in turn causes activation of cyclin D1 and MEC proliferation (Hovey et al., 2003; Brisken et al., 2002). Though details still need to be clarified and confirmed, it has been revealed that both PRL and its mediators are required for normal alveologenesis and roles for these individual mediators are starting to be uncovered.

Both P and PRL are necessary for the massive proliferation and differentiation seen in the beginning stages of pregnancy to prepare for mass production of milk components, but their roles shift in the later stages of pregnancy prior to parturition and lactation. Progesterone suppresses milk secretion during this time while PRL leads to STAT5 induction of milk component expression in the differentiating cells. The removal of the progesterone signal along with high levels of PRL is what eventually allows for lactation (Neville et al., 2002; Anderson et al., 2007). During the withdrawal of P in the MECs, colostrum is produced with the transport of immunoglobulin and transferrin in addition to other milk components prior to parturition (Neville et al., 2001).

1.2.2.4 Lactation

Due to the influences of PRL and P during pregnancy, the mammary tree is complete at the time of initiation of lactation. Primary ducts and side branches have filled the fat pad, which now contains large, fully developed alveoli that are already producing milk components (Anderson et al., 2007). Initiation of lactation occurs around the time of parturition in response to dropping levels of P. In rodents and cows, the level of P sharply decreases and PRL levels spike just prior to parturition, so they are fully lactating when the offspring are born. The level of P in humans does not fall until after parturition, at which point it drops 10-fold in 4 days. The concentration of PRL is already high in humans at parturition as it continuously increases during pregnancy (Neville et al., 2001; Anderson et al., 2007). The placenta and corpus luteum are no longer producing P; the main players from the endocrine system are now the pituitary hormones PRL and oxytocin with the help of a few others to drive metabolic processes (Fendrick et al., 1998).

When stimulated during suckling, the nerve endings at the tip of the nipple send signals to the central nervous system to order the release of hormones including PRL and oxytocin from the pituitary gland (Koyama et al., 2013). Prolactin acts on the secretory MEC to stimulate continued milk production and secretion while the oxytocin causes contraction of the myoepithelium leading to the transport of the milk from the alveoli, travel through the ducts, and ejection to the offspring. As long as suckling and initiation of this mechanism is continued, oxytocin continues to eject the milk and PRL elicits its production. Though PRL is thought to be the main hormone of milk production, other hormones have been implicated as having a role in the process.

There has been some speculation about the roles of insulin, cortisol, thyroid hormone and growth hormone (GH) in governing lactation alongside the influence of PRL. Though initially thought to help the mammary gland increase milk production directly, insulin levels were found to be low in most lactating animals, which is thought to aid the mammary gland in a different way – by discontinuing normal storage mechanisms and help mobilize energy stores from other parts of the body to provide the mammary gland nutrients for milk production (Vernon, 1989; Neville et al., 2002). Although low levels appear to help mammary milk production, the presence of some insulin is necessary for transcription of milk proteins (Trott et al., 2012). Cortisol appears to be elevated in dairy cattle and rodents during lactation, but often at low levels in lactating women. Since it is responsible for maintaining blood glucose it has been speculated that this is the case due to energy store depletion and nutrient availability during lactation of the different species rather than a direct effect on the mammary gland (Neville et al., 2002; Feng et al., 1995; Butte et al., 1999). Thyroid hormone has been shown to be important for the MEC responsiveness to PRL and GH, but has not gained much attention in the recent years (Capuco et al., 1999). Receptors for GH have been found on the mammary epithelium and in the stroma in both types of animal, so it has the ability to act directly on the MEC and can influence milk production by causing expression of IGF-1 both in the mammary stroma and the liver, which can then act directly on the MEC. Growth hormone has proven to be more influential on ruminant lactation after the beginning stages than it appears to be in rodents (Vernon, 1989). Exogenous GH increases milk yield in cattle, which has led to its controversial use in the dairy industry for increased product (Bauman and Vernon, 1993; Neville et al., 2002). Inhibition of GH in rodents has been shown to

slightly reduce milk yield (Flint and Vernon, 1998), but even though GH and PRL have very similar receptors and elicit similar effects, PRL is still considered the most influential of the hormones in milk production. Prolactin inhibition at the onset of lactation greatly affects milk production in many species, including cattle (Karg and Schams, 1974; Neville et al., 2002). Though the detailed intracellular signaling involved in the influence of PRL on milk production is not well known, its role in milk protein synthesis, production of lipogenic and glycolytic enzymes, transporter expression, and lipid droplet formation has been shown and is being continuously investigated (Rudolph et al., 2011). Phosphorylated STAT5 and Akt1 are thought to play major roles in the transcription of many of the genes involved in synthesis of milk components (Anderson et al., 2007; Liu et al., 1997) and recent links between STAT5 and Akt1 have been found, though the specifics of these links are variable (Chen et al., 2010; Creamer et al., 2010; Chen et al., 2012). One link between the two showed that STAT5 induced expression of Akt1, which aided in cell survival and delayed involution during lactation. The specifics of all the intracellular mechanisms behind inducing the production of milk are still being investigated, but the components of milk and the metabolism of these products are relatively well understood.

1.2.2.5 Synthesis of milk components

Milk contains all of the necessary nutrients for the offspring to survive and grow. These include carbohydrate primarily in the form of lactose, lipids primarily in the form of acylglycerol, a variety of proteins including the major milk proteins, caseins and whey proteins, vitamins and minerals, and water. There are a huge variety of molecules present in milk so only the major constituents will be further discussed. All of these components are synthesized and secreted, or transported from the blood to the milk, by the secretory MEC lining the lumen of the alveoli. Specifics of the synthesis and secretion of these components including a mention of any known hormonal influence on their regulation will be covered in the following text.

1.2.2.5.1 Lactose synthesis

Lactose is the primary carbohydrate found in the milk of most mammals and is the major osmotic constituent responsible for drawing water into the final product of milk. Water makes up the majority of milk, with all other components being dissolved or suspended in it. Lactose concentration is rather constant within a species and relatively constant across species when compared with other milk components. In humans, cattle and mice lactose makes up approximately 7%, 5%, and 2.5% of milk, respectively (Jenness, 1974; Gors et al., 2009; Johnson et al., 2001). Synthesis of lactose is unique to the secretory MEC and is carried out by the enzyme lactose synthetase. Lactose synthetase is made up of galactosyltransferase and its cofactor, α-lactalbumin, which is also a secreted milk protein. Expression of both of the enzyme's subunits can be induced by PRL and there is evidence that insulin is necessary for expression of α-lactalbumin (Palmiter, 1969;

Turkington et al., 1968; Rosen et al., 1999). This enzyme catalyzes the reaction between UDP-galactose and glucose to create lactose and UDP (Kuhn et al., 1980). As even UDP-galactose is derived primarily from glucose in the secretory MEC, glucose is necessary and in high demand for lactose synthesis (Faulkner and Peaker, 1987). A glucose transporter independent of insulin regulation is responsible for maintaining a high glucose concentration within the cell as well as allowing glucose into the Golgi compartment to support lactose synthesis; this transporter is GLUT1 and its expression is increased under the influence of PRL during lactation. Its location on the Golgi membrane is unique to the secretory MEC (Nemeth et al., 2000; Anderson et al., 2007). Glucose is in high demand in the secretory MEC during lactation for the synthesis of milk components other than just lactose.

1.2.2.5.2 Lipid synthesis

Triacylglycerol (TAG) makes up 98% of the fat in milk (Anderson et al., 2007; Jenness, 1988), while total fat in milk varies among species ranging from nearly 0% up to around 50%. Humans and cattle produce milk with an average of 4% fat while mice produce milk with approximately 20% fat content (Neville and Picciano, 1997; Gors et al., 2009; Johnson et al., 2001). Milk fat content can also largely vary within species dependent on body condition, diet, and stage of lactation (Neville and Picciano, 1997). Triacylglycerol is made up of a glycerol backbone esterified to three fatty acids (FA). Fatty acids are either transported into the cell from adipose stores, supplied by diet, or synthesized *de novo* in the MEC from glucose, amino acids, or in the case of ruminants, acetate. Composition and source of FA can also vary with species, diet and body condition. Long

chain FA are typically delivered to the cell by albumin carriers and lipoproteins or in the form of TAG from the diet or lipid stores of the body via the blood in which it is broken down by lipoprotein lipase into FA and glycerol before being taken up into the MEC. Medium chain FA are synthesized in the MEC from glucose in nonruminants, or acetate in ruminants, or amino acids in both and their levels are reduced if high dietary fat is present (Neville and Picciano, 1997; Anderson et al., 2007). Unlike in nonruminants, short chain FA are synthesized *de novo* and found in ruminant milk (Knight and Beever, 1994; Neville and Picciano, 1997). Glycerol-3-phosphate is used as the glycerol backbone to which the FA attach and form TAG. It is either transported into the cell from the blood in the form of glycerol, which is then phosphorylated, or it is synthesized from dihydroxyacetone phosphate (DHAP), which originates from glucose (Salway, 1994; Anderson et al., 2007). Glucose also contributes to NADPH production. Fatty acid synthesis requires NADPH as a reducing agent whether they are being synthesized from acetate or glucose precursors. NADPH can be generated through different pathways that vary with species. Non-ruminants can generate NADPH through the malate/pyruvate pathway, the isocitrate pathway, or through the pentose phosphate shunt (PPS). The conversion of malate to pyruvate by malate dehydrogenase generates NADPH and allows pyruvate to re-enter the TCA cycle. Ruminants lack this ability as they have very low malate dehydrogenase and citrate lyase activity. Ruminants do have high isocitrate dehydrogenase activity and NADPH can also be generated by the conversion of isocitrate to α-ketoglutarate by this enzyme (Ingle et al., 1972). The last method of NADPH generation is by glucose shuttling into the PPS. It enters the PPS as glucose-6-phosphate (G6P) and can then re-enter the glycolytic pathway as G6P after passing through the

shunt (Salway, 1994). Prolactin inhibition causes levels of lipoprotein lipase, lipogenic and glycolytic enzymes and enzymes involved in the PPS to fall significantly, indicating that this hormone plays a regulatory role in lipid synthesis in the MEC (Rudolph et al., 2011).

1.2.2.5.3 Protein synthesis

Protein synthesis occurring in the MEC must support both milk protein generation and generation of proteins responsible for maintaining the cell and for producing and transporting other milk components. Protein constitutes approximately 0.9%, 3.5%, and 12% of milk in humans, cattle, and mice, respectively (Jensen, 1995; Lonnerdal, 1985; Gors et al., 2009; Boumahrou et al., 2009). Major milk proteins serve as carriers for necessary vitamins and minerals, act as antimicrobials, anti-inflammatories, antioxidants and specific immune protection, and serve as peptide and amino acid sources to the growing offspring. Hormones and growth factors are also present in the milk, which are thought to have possible roles in the neonate as well (Thompson et al., 2009). There are two classifications of protein in milk: the caseins and the whey proteins. Caseins are proteins that are insoluble in acidic pH range, while whey proteins are the proteins left in the milk when caseins are removed by acidification. Caseins primarily exist in milk in the form of micelles, which are structures formed with different major casein proteins stabilized by kappa-casein and organic and inorganic ions, such as calcium (Jenness, 1974; Lonnerdal, 1985). Composition of these casein micelles and the whey protein portion of milk vary among species. The only major casein in human milk is beta-casein, which is almost entirely bound in micelles (Lonnderdal, 1985). The major casein in

bovine milk is αs1-casein, though it also contains a large amount of beta-casein and smaller amounts of αs2- and gamma-casein. Fewer of the caseins in cow milk are bound in micelles than in human milk (Jensen, 1995). Mouse milk contains large amounts of both αs1- and beta- casein with smaller amounts of gamma-casein (Boumahrou et al., 2009). All three contain smaller amounts of kappa-casein, which helps the others form micelles in the milk. The whey portion of milk consists of different proportions of α-lactalbumin, lactoferrin, beta-lactoglobulin, immunoglobulins, serum albumin, whey acidic protein, enzymes, hormones and growth factors. Beta-lactoglobulin is primarily found in cow milk, while whey acidic protein is primarily found in mouse milk. Immunoglobulin and serum albumin are not made in the MEC, but transported from the blood into the milk by these cells instead (Boumahrou et al., 2009; Kinsella and Whitehead, 1989; Lonnderdal, 1985).

In addition to the secreted milk proteins, the cell is generating the proteins involved in their production and the production of other milk components and cell maintenance requirements. To support the high protein synthesis and processing demand, MEC prepare for lactation by synthesizing more Golgi and rough endoplasmic reticulum (Hollman, 1974), which can then aid in the production of the necessary proteins for milk production. Among these proteins being produced are receptors for hormones and growth factors secreted by other cells as well as components of cell signaling initiated by activation of the receptors that cue the cell to produce the proteins needed for the jobs it has to perform. Milk protein production as well as transport of immunoglobulin is largely regulated by PRL and phosphorylation of PRLR's downstream transcription factor, STAT5. As previously mentioned, expression of enzymes involved in lactose synthesis

and glycolytic and lipogenic pathways are also regulated by this hormone (Anderson et al., 2007). Seen by the diverse roles of proteins in the MEC, the proper regulation of protein synthesis is important for every aspect of milk production and secretion.

1.2.2.6 Involution

After completion of functional development and the duration of lactation, the offspring no longer require milk for nourishment and the gland undergoes involution, returning to a near virgin state. This process has been most studied in mice and the following information will stem from studies done in this species. Involution naturally occurs gradually as offspring slowly shift to other food sources, though most studies investigate the effects of abrupt weaning during mid-lactation in order to investigate the initiation of different events occurring during involution (Richert et al., 2000). Consequently, timing mentioned here is based off of abrupt weaning and involution while these events most likely occur naturally in a similar manner but spread over the time of weaning.

Involution is a two-step process characterized by an early and reversible phase of MEC apoptosis and shedding, followed by an irreversible stage of ECM and gland remodeling. The initiation of involution is caused by milk stasis in the alveolar lumen and can be reversed by suckling until remodeling has begun. In mice, the first stage of reversibility lasts approximately 48 hours (Watson, 2006). When milk builds up in the lumen of the alveoli, the MEC appear flattened due to the pressure of the building milk volume in the extended alveoli (Richert et al., 2000) and this milk stasis induces expression of many different factors implicated in the apoptotic pathways of involution. Milk accumulation induces the expression of leukemia inhibitory factor (Lif), which appears in the alveolar lumen and activates Lif-receptor (LifR) on the MEC luminal membrane. This subsequently activates the Jak/STAT pathway, leading to the phosphorylation and activation of STAT3. While STAT5 is important in development of the functional mammary gland, STAT3 is essential in its involution after lactation (Watson et al., 2006;

Stein et al., 2007). STAT3 induces expression of several important players in the onset of involution such as C/EBPδ, IGFBP5 and regulatory subunits of PI3-kinase which are involved in the anti-inflammatory pathway and inhibition of Akt activation, which would otherwise aid in cell survival (Stein et al., 2007). Vitamin D receptor and the TNF/death receptor pathway take a more direct approach. While they inhibit Akt, they also induce caspase activity, which directly leads to apoptosis. At this stage of involution caspase acitivty is only seen in cells shed into the lumen (Stein et al., 2007; Watson et al., 2006).

During the beginning phase of involution the only difference apparent in the gland is the increased shedding of MEC, but the irreversible second stage is characterized by major remodeling of the epithelium and stroma. Alveolar structures begin to collapse due to detachment from the ECM and caspase activity is now seen in MEC still attached to the luminal wall. Matrix metalloproteases (MMPs) are essential to this second stage of tissue remodeling. They are responsible for the remodeling of the ECM and detachment from the MEC layer. They also cleave plasminogen to plasmin, which is required for normal stromal remodeling and adipocyte repopulation (Watson, 2006). When the alveoli collapse, adipocytes begin accumulating lipid and filling in the open space of the stroma. Infiltrating immune cells and surviving MEC perform clearance of dead cells and debris. Resident MEC engulf milk fat globules, casein micelles, and apoptotic MEC and have the ability to release the same key cytokines in response as the professional phagocytes. The professional phagocytes also play a role in the clean up; macrophages and neutrophils engulf milk components and phagocytize apoptotic cells (Atabai et al., 2007; Watson et al., 2006). The rearranging and clearance continues until the gland has returned to a mature, quiescent state that nearly mimics an adult virgin gland, which in mice is

approximately 10-15 days post weaning (Richert et al, 2000; Lund et al., 1996; Lascelles and Lee, 1978).

1.2.2.7 Quiescence

After involution is complete, the mammary gland enters a period of quiescence until pregnancy initiates another surge in growth and development. Although the post-involution gland appears very similar to the mature virgin gland, there are persisting differences. First, the morphological appearance of the gland varies slightly from the virgin state. In humans and mice the complexity of the glandular tissue in the parous gland is higher than in the nulliparous gland (Richert et al, 2000; Daniel and Silberstein, 1987). In addition to slight morphological differences, persisting molecular differences exist. In a study by D'Cruz and colleagues (2002) the lasting effects of parity on mammary gland gene expression was analyzed by microarray. Among the genes that were down-regulated with parity were growth factors such as amphiregulin and IGF-1, which promote cell growth and survival. Among the much longer list of up-regulated genes were transcripts indicating a higher state of differentiation, such as all major milk proteins, as well as immunoglobulin and factors involved in TGF-β signaling. These differentially expressed transcripts indicate a higher order of differentiation as well as more of a presence of immune cells and increased incidence of growth inhibition. If the proteins of these transcripts are also expressed, this may help explain the well-documented protection of parity from breast cancer (MacMahon et al., 1970; Britt et al., 2007).

1.2.3 Breast cancer incidence

Effect of parity on breast cancer risk has been recognized and documented since the early 1900s (Lane-Claypon, 1926; Wainwright, 1931). Both reduced and increased risks are associated with different parity conditions. A transient increase in the risk of breast cancer immediately following a pregnancy has been documented, which is also decreased with increasing parity (Lambe et al., 1994). Risk of breast cancer development was found to be reduced by 10% with each additional full term pregnancy (Lambe et al., 1996). Age of first parity also has a large influence on risk. Women who are older than thirty at their first full term pregnancy have a higher risk of developing breast cancer than nulliparous women (Kelsey et al., 1993). While late age of first parity has a detrimental effect, early age at first parity decreases risk of breast cancer development. The reduced risk of breast cancer due to early age at first full term pregnancy became well known when MacMahon and colleages performed a study in 7 different areas around the world and recognized and stressed the importance of this factor (1970). Since then, studies in rodents and humans have confirmed this protective factor of early pregnancy and have found that steroid receptor positive breast cancer risk is decreased by early childbirth (Britt et al., 2007; Ma et al., 2006). Several studies have focused on identification of the molecular factors responsible for this phenomenon (D'Cruz et al., 2002; Ginger et al., 2001; Balogh et al., 2006), and some potential molecules and processes have been discovered, but too few reports and too many potential factors make concrete conclusions impossible. The studies mentioned also all looked at persisting transcripts, while persisting protein expression might provide a more focused list of potential future investigations. Continued

comparisons between the different states of the gland offer increased potential of unveiling persisting factors that might influence this difference seen in breast cancer risk.

1.3 Bioinformatics

The expanding field of bioinformatics and the increasing importance for interdisciplinary studies has been well documented and is only growing more apparent (Wolkenhauer and Hofmeyr, 2013; NRC, 2010; Howe et al., 2008). Few experimentation techniques in biological research remain that do not utilize computer software for some aspect of the study and computers are necessary for storing, organizing and performing the analysis of data. With the emergence of systems biology and "–omic" technologies, computer analysis and data sharing has become fundamental to all progress. While individual fields of study are becoming narrower in their focus and splitting into more divisions, these divisions are becoming more difficult to assign under a larger category; that is, the lines between fields of study are becoming blurred. It is becoming more important to have a larger variety of knowledge in order to understand the details of these smaller topics. To go along with these trends, individualized computer programming is being generated to fit the needs of the specific fields of study. At the 2013 Plant and Animal Genome conference in San Diego, an unofficial survey revealed that approximately 55 abstracts were presented detailing computer programs designed to aid in -omic data management. Among them, topics addressed included genome annotation, genotype and phenotype association, transcriptome assessment, and regulation of gene expression. With the narrowing focus of study topics and the growing amount of data that can be generated, this observation did not come as a surprise.

Need for this cross-discipline has been recognized and attempts are being made to advance its progress. Classes introducing undergraduates to bioinformatics tools have appeared in many universities. Barbara May reported on the success of a class aimed at

this goal. Students learned to analyze genomes and annotate genes, giving mostly positive feedback (May, 2013). May is the far from the first to begin these types of studies; in fact, it is now very common to even find degrees available involving bioinformatics in universities. A survey was originally compiled in 2002 for a meeting of the American Society of Information Science and Technology that included a list of universities that offered bioinformatics related programs and degrees, and is continuously updated by Bradley Hemminger at the School of Information and Library Services at the University of North Carolina at Chapel Hill. In 2002 the list from the survey included 27 different programs and by 2012 the list had increased to 87 different programs; that is 60 new programs in ten years (SILS, 2013). The programs offer degrees related to bioinformatics ranging from certificate programs and minors to BS, MS and PhD degrees. The degree programs are based in departments ranging from biology to computer science to mathematics departments. Some universities have their own department dedicated to bioinformatics, and some are interdepartmental or interinstitutional programs.

Huge progress has been made in the recent years toward interdisciplinary studies, but many of the programs are still new and continued advancement and improvement of these programs is necessary for the advancement of the biological studies as the data sets being produced are overwhelming without the help of advancing technology.

1.4 Summary

Development of the mammary gland through the life of the female mammal is under the influence of hormonal and intracellular signals, many of which are yet to be discovered. These signals allow for the differentiation of the MECs into entities able to produce and secrete the components of milk. In addition to the differentiation into milk producing phenotypes, the cells also regress to a quiescent state during times of non-pregnancy. Understanding of the events behind these developmental stages is of interest to several different populations because of the potential influence on dairy production, breast feeding, and breast cancer research.

Investigation into the development of the mammary gland is often performed by utilizing large scale investigative methods such as microarray and proteomics. This type of analysis can be made easier with the help of computer applications and collaboration with technological fields. Specific computer software for a lab's individual purposes can relieve many of the difficulties that come with analyzing large data sets. Emerging interdisciplinary study programs that allow students to learn the technological side and biological side of this type of research will be of much value to future generations of scientists.

CHAPTER 2 – Proteomic Analysis of HC11 Differentiation

2.1 Introduction

The role of the differentiated mammary epithelial cell (MEC) during lactation is to produce and transport all of the essential nutrients for the organism's offspring including lactose, fatty acids and triacylglycerol, amino acids and proteins, vitamins, minerals, and several other important components needed for the growing infant (Mepham, 1983). The nutrients available to the MEC, metabolic capabilities of the organism, and nutritional needs of the young determine the components used and secreted in the milk. Because of these determining factors some differences in molecular processing and the metabolic pathways leading to the end products found in milk would be expected among species and individuals, but despite the differences expected, the overall story should remain similar due to the responsibility of the functional mammary gland across all mammals.

Mammary differentiation is studied through *in vitro* cell culture experiments, tissue explants, and whole tissue samples (Desrivieres et al., 2003; Collier et al., 1977; D'Cruz et al., 2002). Each method of study has advantages and disadvantages. Cell culture allows for study of individual cell types, but removes the natural environment of the cells and cell lines can misrepresent their primary cell counterparts. Explants and whole tissue samples may provide more of a natural environment for the study of the gland, but information is obtained for the gland as a whole, and information on individual cell types is lost. A recent investigation in the Peterson lab was done using isolated primary MEC from pregnant and lactating mice, which allows development in the natural environment and the investigation of an individual cell type (Strand, 2012). Comparison of this data to that of the HC11 cell line differentiated in culture would provide valuable information.

This experiment analyzes the differing protein profiles between the undifferentiated and differentiated states of the murine HC11 MEC line and compares this analysis with different models of MEC differentiation. While the process of differentiation cannot be fully understood with a just snapshot of the proteome on either side of the main event, this analysis distinguishes aspects of MEC differentiation for which cell lines may or may not be appropriate models and indicates the importance of comparative analysis among species and models to reveal commonalities, which will be important for future attempts at identifying central regulatory networks of MEC development.

2.2 Methods

2.2.1 Cell culture

HC11 murine mammary epithelial cells (Ball et al., 1988) were grown to confluence in RPMI 1640 media supplemented with 10% fetal bovine serum (FBS; Innovative Research, Novi, MI), 1% penicillin/streptomycin (PS), 1% amphotericin B, 5 μg/ml insulin, and 10 ng/ml epidermal growth factor (EGF). Half of the cells (n=3) were taken for protein extraction at the point of confluence. Differentiation was then induced in the remaining cells by incubating with EGF-free growth media for 48 hours followed by 4 days incubation in RPMI 1640 with 10% FBS, 1% PS, 1% amphotericin B, 5 μg/ml insulin, 1μM dexamethasone, and 1μg/ml prolactin. The remaining cells (n=3) were used for protein extraction after day 4 when domes were present in the cells, indicating a differentiated state.

2.2.2 Protein extraction

Three plates of cells were pooled for each protein extraction sample leading to 3 samples per treatment (n=3). Media was aspirated from plates and cells were transferred to microcentrifuge tubes and rinsed with PBS. Cells were then lysed by sonication in homogenization buffer (7 M urea, 2 M thiourea, 40 mM tris base, 1% ASB-14, 40 mM DTT, 0.5% ampholyte IPG, 0.001% bromophenol blue). Lysate was separated by centrifugation for 30 minutes at 10,400 x g and 4°C and supernatant containing isolated soluble proteins was transferred to a new tube.

Protein was precipitated with 10% trichloroacetic acid in acetone overnight at -20°C. Protein was rinsed with 100% acetone and allowed to dry. Protein was solubilized overnight at 4°C in rehydration buffer containing 7 M urea, 2 M thiourea, 2% CHAPS, 2% nonidet P-40, 100 mM DTE, 0.5% ampholyte IPG, and 0.002% bromophenol blue. After centrifugation for 15 min at max speed and 4°C, supernatant containing solubilized protein was transferred to a clean tube and stored at -80°C.

Protein was quantified using the 2-D Quant Kit (GE Healthcare Life Sciences, Pittsburgh, PA).

2.2.3 Two-dimensional gel electrophoresis (2DGE)

All equipment and materials used for 2DGE were purchased from Bio-Rad (Hercules, CA) unless otherwise stated. All buffer reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated.

Immobilized pH gradient strips (11 cm, pH 3 -10) were actively rehydrated with the rehydration buffer containing the protein samples for 12 hours at 50 V. Isoelectric focusing (IEF) for the first dimension of separation was then performed at ~8,000 V and 20°C for 35,000 Volt hours. Active rehydration and IEF were performed using the Protean IEF Cell. Strips were stored at -80°C until subjected to the second dimension.

For the second dimension, IPG strips containing protein were incubated with equilibration buffer (375 mM tris-HCl (pH 8.8), 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue) containing 10 mg/ml DTT on a rotator for 15 min at room temperature followed by incubation with equilibration buffer containing 25 mg/ml iodoacetamide for 15 min. Proteins were then separated by molecular mass using 11 cm 10% polyacrylamide Criterion tris-HCl gels using the Criterion Dodeca Cell at 200 V, allowing all gels to be run simultaneously. Samples were run in duplicate and proteins in all gels were stained overnight with colloidal Coomassie Blue G-250 and de-stained with Type I DI water.

2.2.4 Gel analysis, spot picking and trypsin digestion

Stained gels were scanned using an Epson 1280 transparency scanner (Epson, Long Beach, CA, USA). Scanned gel images were processed and analyzed by Delta 2D (version 3.6, Decodon, Greifswald, Germany). Spots boundaries were defined and gels were overlaid and fitted to align corresponding spots across gels. Differentially expressed protein spots were identified using a *t*-test performed according to a null distribution that was generated with 1000 permutations in order to account for unequal variance and non-normal distribution of data.

Protein spots that differed in abundance due to treatment were excised using a manual 1.5 mm tissue puncher (Beecher Instruments, Prairie, WI) and stored at -80°C in 0.5 ml microcentrifuge tubes until further processing. Gel plugs containing individual protein spots were destained twice by incubation for 30 min at room temperature on a shaker with destaining buffer (25 mM ammonium bicarbonate in 50% acetonitrile), dehydrated with 100% acetonitrile, and digested overnight with trypsin solution (11 µg/µl MS-grade porcine trypsin gold (Promega, Madison, WI) in 40mM ammonium bicarbonate/10% acetonitrile) at 37°C. Digested proteins were eluted with analyte solution (0.1% trifluoroacetic acid (TFA)/acetonitrile 2:1) for 30 min on a shaker at room temperature, repeated twice. Samples were concentrated using a SpeedVac (Thermo Fisher Scientific, Waltham, MA) at 45°C, resuspended in 6 µl of matrix solution (0.2 mg/ml α-cyano-4-hydroxycinnamic acid in acetonitrile) and plated on an Anchorchip target plate (Bruker Daltonics Inc., Billerica, MA). Plated protein spots were washed with 0.1% TFA and recrystallized with acetone/ethanol/0.1% TFA (6:3:1).

2.2.5 Mass spectrometry and protein identification

Peptide mass fingerprints (PMFs) were obtained using a matrix-assisted laser desorption ionization tandem time-of-flight (MALDI TOF/TOF) mass spectrometer (Ultraflex II; Bruker Daltonics Inc., Billerica, MA). Trypsin was used for internal mass calibration. PMFs were analyzed using MASCOT server launched from BioTools software (Bruker Daltonics, Billerica, MA) against the NCBI database. PMF were further analyzed using MS/MS spectra using five to ten of the largest peaks per sample (excluding keratin and trypsin). Spectra were internally calibrated and processed using FlexAnalysis software (Bruker Daltonics, Billerica, MA). PMF and MS/MS spectra were combined and queried as described for PMF spectra analysis using the MS/MS spectra.

2.3 Results

All gels containing both undifferentiated and differentiated HC11 cell protein extracts were scanned and overlaid, creating one fused image for use in spot detection. Delta2D analysis of the fused gel image detected a total of 337 individual protein spots (Figure 1). Of the detected spots, 98 of these spots differed in abundance between the two treatments ($P < 0.02$), and 90 of these differentially expressed spots were identified when using the MASCOT server and matching the spectra to the NCBI Database (Table 1). Of the 90 identified spots, 41 were up regulated and 49 were down regulated in the differentiated HC11 cells when compared to the undifferentiated cells.

Functional groups for the majority of the identified differentially expressed proteins include energy metabolism (24 spots total; 18 up regulated/6 down regulated), cytoskeletal dynamics (16 spots total; 6 up regulated/10 down regulated), transcriptional regulation and RNA processing (11 spots total; 8 up regulated/3 down regulated), and protein folding and processing (27 spots; 6 up regulated/21 down regulated).

Figure 1. *Fused gel image of HC11 protein spots*. Protein was extracted from undifferentiated and differentiated HC11 cells and protein was subjected to 2-dimensional gel electrophoresis, stained with Coomassie blue, and analyzed by Delta 2D. The image represents all gels fused into one image. Labeled spots were differentially expressed between treatments ($P < 0.02$). Numbers correspond to "Spot ID" in Table 1.

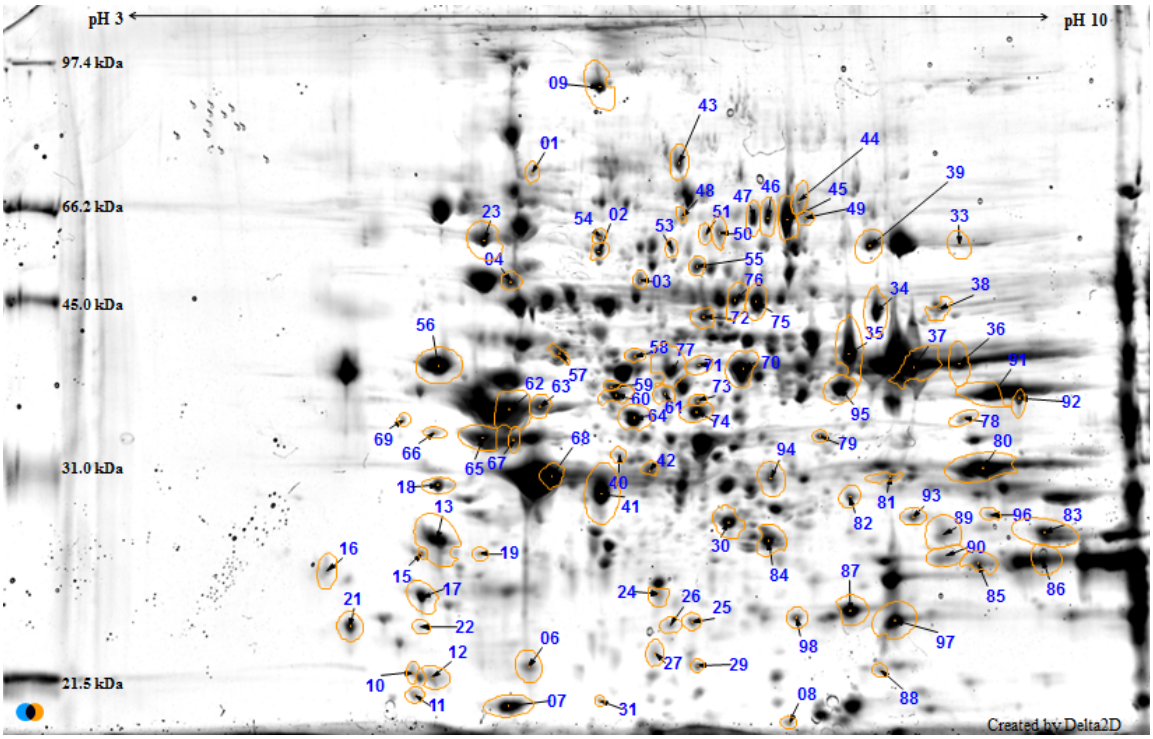

Figure 2. *Heat map of differentially expressed HC11 protein spots*. Created by hierarchical clustering using Pearson's correlation of HC11 protein spots that were differentially expressed between the undifferentiated and differentiated states (P<0.02). Blue represents lower than average protein abundance, while orange represents higher than average protein abundance for that particular spot. Rows represent individual spots and the numbers correspond to "Spot ID" in Table 1. Columns represent individual gels. Differentiated treatment gels are on the left and undifferentiated treatment gels are on the right.

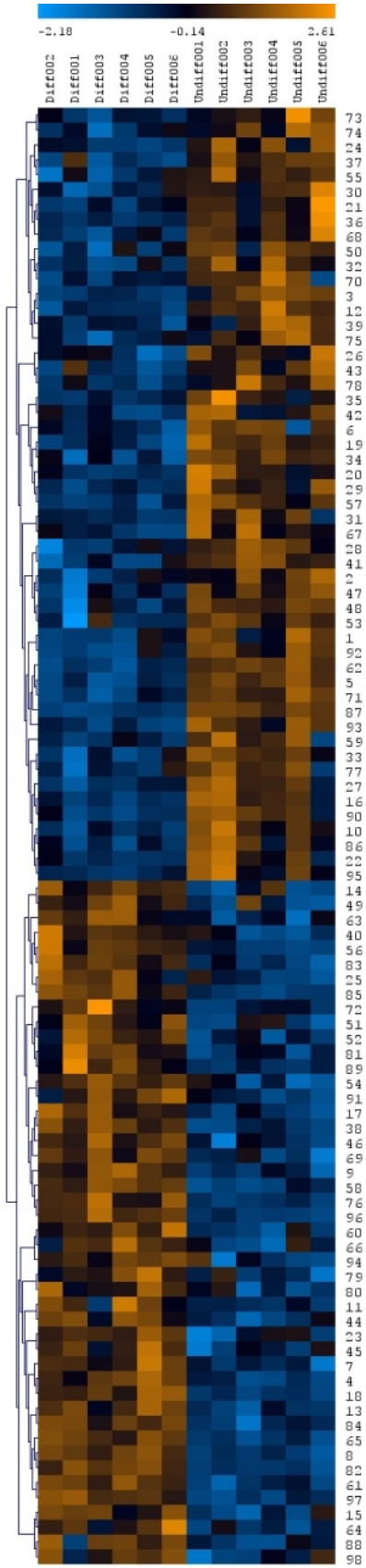

Figure 3. *Proposed energy metabolism pathways in differentiated HC11 cells*. Green represents up regulation and red represents down regulation in differentiated cells ($P < 0.02$). The protein expression profile indicated shifts in pathway flux to spare glucose for lactose synthesis, support energy generation from amino acids and redirect amino acid carbon skeletons, oxaloacetate and acetyl CoA out of the mitochondria for use in triacylglycerol synthesis in the cytosol.

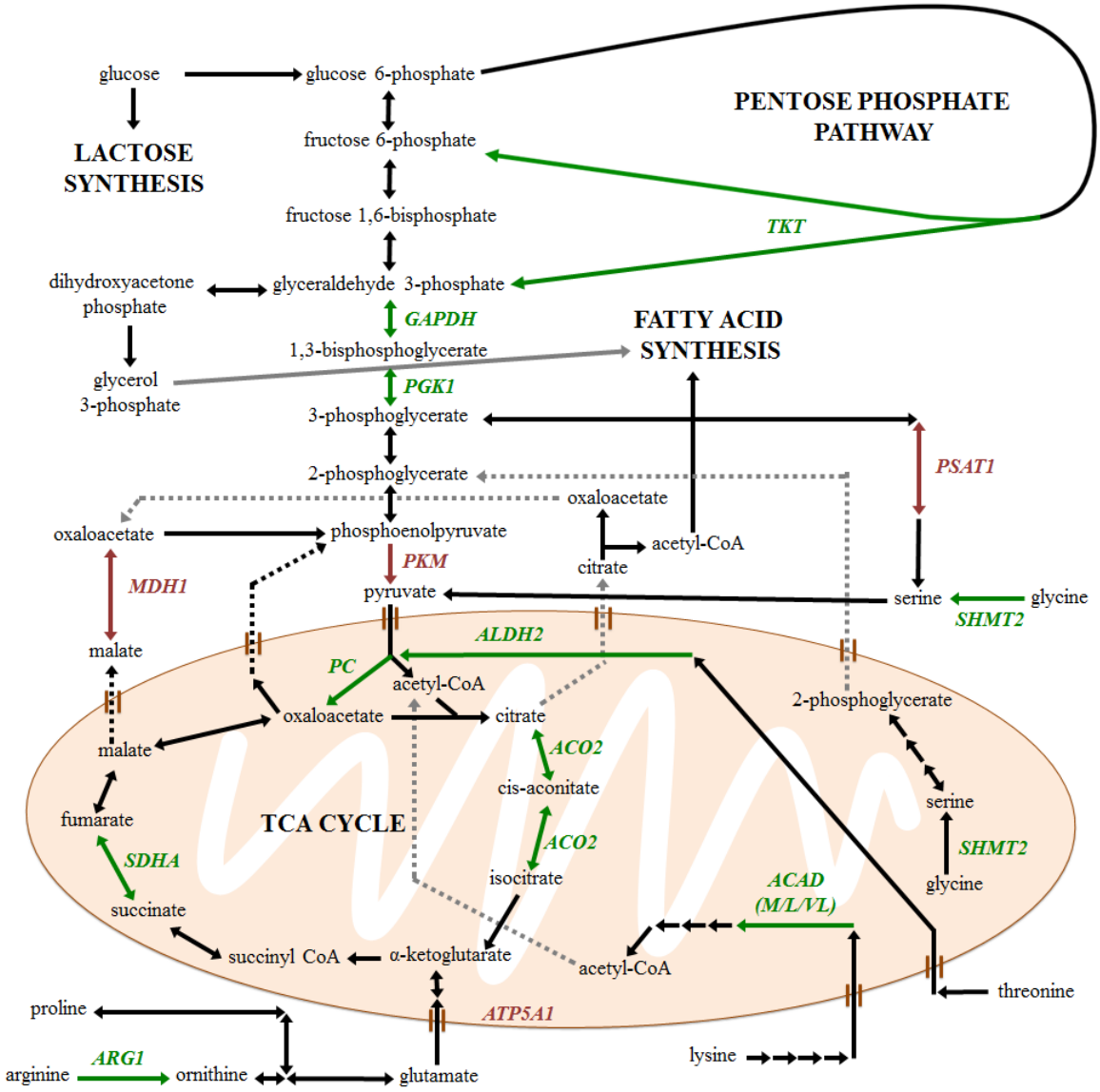

Table 1. *List of identified protein spots that were differentially expressed between undifferentiated and differentiated HC11 cells* ($P < 0.02$). Under "Direction of Regulation," "up" indicates an increased abundance of the specified protein, while "down" means the decreased abundance of the specified protein in the differentiated cells relative to the undifferentiated cells. "Fold change" indicates magnitude of abundance change and was calculated by dividing mean spot abundance of differentiated gels by mean spot abundance of undifferentiated gels.

Spot ID	Protein Name	Gene Name	Direction of Regulation	Fold change
	angiogenesis & differentiation			
40	angiopoietin-2	Angpt2	up	1.90
	cell signaling			
21	complement component 1 Q subcomponent-binding protein, mitochondrial	C1qbp	down	0.63
97	guanine nucleotide-binding protein subunit beta-2-like 1	Gnb2l1	down	0.75
22	high affinity nerve growth factor receptor precursor	Ntrk1	down	0.47
12	14-3-3 gamma protein	Ywhag	down	0.69
10	14-3-3 protein theta	Ywhaq	down	0.40
	cytoskeleton			
68	beta-actin	Actb	up	1.21
54	gelsolin, cytosolic	Gsn	down	0.24
2	gelsolin, cytosolic	Gsn	up	2.67
9	integrin beta-4, partial	Itgb4	down	0.32
63	keratin, type I cytoskeletal 14	Krt14	down	0.48
66	keratin, type I cytoskeletal 15	Krt15	up	2.55
65	keratin, type I cytoskeletal 17	Krt17	down	0.48
67	keratin, type I cytoskeletal 17	Krt17	down	0.60
70	keratin, type II cytoskeletal 5	Krt5	down	0.49
37	keratin, type II cytoskeletal 5	Krt5	down	0.49
71	keratin, type II cytoskeletal 5	Krt5	down	0.66
77	keratin, type II cytoskeletal 5	Krt5	down	0.74
59	keratin, type II cytoskeletal 7	Krt7	down	0.49
75	prelamin-A/C isoform A	Lmna	up	1.33
76	prelamin-A/C isoform A	Lmna	up	1.34
60	septin-8, partial	Sept8	up	2.45
	DNA replication & cell division			
53	DNA replication licensing factor	Mcm7	down	0.64
17	proliferating cell nuclear antigen	Pcna	down	0.14
73	ruvB-like 1	Ruvbl1	down	0.48
	ion transport			
55	mitochondrial inner membrane protein, partial	Immt	up	1.30
88	voltage-dependent anion-selective channel protein 2	Vdac2	down	0.27

Spot ID	Protein Name	Gene Name	Direction of Regulation	Fold change
	metabolism			
94	long-chain specific acyl-CoA dehydrogenase, mitochondrial precursor	Acadl	up	1.53
81	medium-chain specific acyl-CoA dehydrogenase, mitochondrial precursor	Acadm	up	1.86
38	very-long-chain acyl-CoA dehydrogenase	Acadvl	up	1.70
39	aconitate hydratase, mitochondrial precursor	Aco2	up	1.45
74	aldehyde dehydrogenase, mitochondrial precursor	Aldh2	up	2.19
84	arginase-1	Arg1	up	2.10
30	arginase-1	Arg1	up	2.17
92	ATP synthase subunit alpha, mitochondrial precursor	Atp5a1	down	0.42
98	electron transfer flavoprotein subunit alpha, mitochondrial	Etfa	up	2.41
87	glyceraldehyde-3-phosphate dehydrogenase	Gapdh	up	1.28
90	glyceraldehyde-3-phosphate dehydrogenase	Gapdh	up	1.66
26	3-hydroxyisobutyrate dehydrogenase, mitochondrial precursor	Hibadh	up	2.05
24	malate dehydrogenase, cytosolic	Mdh1	down	0.59
25	3-mercaptopyruvate sulfurtransferase	Mpst	up	2.81
44	C-1-tetrahydrofolate synthase, cytoplasmic	Mthfd1	down	0.44
29	omega-amidase	Nit2	down	0.23
95	succinyl-CoA:3-ketoacid coenzyme A transferase 1, mitochondrial precursor	Oxct1	up	1.36
43	pyruvate carboxylase, mitochondrial isoform 1	Pc	up	1.84
80	phosphoglycerate kinase 1	Pgk1	up	1.33
35	pyruvate kinase M	Pkm	down	0.79
96	phosphoserine aminotransferase isoform 1	Psat1	down	0.57
72	succinate dehydrogenase flavoprotein subunit, mitochondrial, partial	Sdha	up	1.77
91	serine hydroxymethyltransferase 2, mitochondrial	Shmt2	up	1.45
34	transketolase	Tkt	up	1.23
	plasma protein			
78	serum albumin precursor	Alb	up	1.61
	protein synthesis & processing			
61	t-complex protein 1 subunit beta	Cct2	down	0.68
36	t-complex protein 1 subunit eta	Cct7	down	0.53
7	cathepsin B	Ctsb	down	0.60
41	cathepsin D precursor	Ctsd	up	1.92
27	cathepsin D precursor	Ctsd	up	2.21
79	dnaJ homolog subfamily A member 1	Dnaja1	down	0.51
19	elongation factor 1-delta	Eef1d	down	0.46
46	elongation factor 2	Eef2	down	0.54
49	elongation factor 2	Eef2	down	0.55
45	elongation factor 2	Eef2	down	0.57
47	elongation factor 2	Eef2	down	0.67
3	glycine - tRNA ligase	Gars	down	0.44
23	endoplasmin	Hsp90b1	down	0.59

Spot ID	Protein Name	Gene Name	Direction of Regulation	Fold change
31	heat shock protein beta-1	Hspb1	down	0.16
57	chaperonin	Hspd1	down	0.42
16	nascent polypeptide-associated complex subunit alpha isoform b	Naca	down	0.37
13	nucleophosmin isoform 1	Npm1	down	0.50
15	nucleophosmin isoform 3	Npm1	down	0.36
56	protein disulfide-isomerase	P4hb	up	1.46
4	protein disulfide-isomerase A4 precursor	Pdia4	up	1.54
50	procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	Plod2	up	1.78
51	procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	Plod2	up	1.90
82	peptidyl-prolyl cis-trans isomerase D	Ppid	down	0.65
8	proteasome subunit alpha type-2	Psma2	down	0.60
11	proteasome subunit alpha type-5	Psma5	down	0.40
18	40S ribosomal protein SA	Rpsa	down	0.71
58	t-complex polypeptide 1A	Tcp1	down	0.60
	transcriptional regulation & processing			
86	heterogeneous nuclear ribonucleoproteins A2/B1 isoform 2	Hnrnpa2b1	up	1.19
85	heterogeneous nuclear ribonucleoproteins A2/B1 isoform 2	Hnrnpa2b1	up	1.74
83	heterogeneous nuclear ribonucleoprotein A3 isoform c	Hnrnpa3	up	1.76
93	heterogeneous nuclear ribonucleoprotein A/B isoform 2	Hnrnpab	down	0.63
64	heterogeneous nuclear ribonucleoprotein H2	Hnrnph2	up	1.32
89	max-like protein X isoform gamma	Mlx	up	2.27
48	tRNA (cytosine(34)-C(5))-methyltransferase	Nsun2	down	0.55
69	ribonuclease/angiogenesis inhibitor	Rnh1	down	0.15
6	TAR DNA-binding protein 43 isoform 1	Tardbp	up	1.58
42	TAR DNA-binding protein 43 isoform 1	Tardbp	up	3.08
1	tripartite motif protein 28	Trim28	up	2.02

2.4 Discussion

In this study, protein expression of the HC11 murine MEC line was compared between its undifferentiated state, used to represent the theoretical state of pregnancy, and its differentiated state, used to represent the theoretical state of lactation. Samples were taken at the proliferating stage and after being treated with a three-hormone cocktail (dexamethasone, insulin, prolactin; DIP), respectively. Treatment with DIP has been shown to yield cells with characteristics of a lactating MEC based on expression of beta-casein and formation of mammospheres (Ball et al., 1988; Blachford et al., 1995; Morrison and Cutler, 2009). The method of analysis used in 2DGE allows the discovery of many protein changes occurring at a point in time before and after differentiation, but it does not provide a complete list of all occurrences leading to the differentiated cell type. This analysis allows the proposal of potential events occurring, which can then be explored and validated through additional experimentation.

Changes in the protein profile of these HC11 cells after differentiation are largely in agreement with a very similar study performed ten years prior (Desrivieres et al., 2003) and appear to indicate shifts appropriate for the switch to a milk producing phenotype. Differential expression of identified proteins involved in energy metabolism indicated shifts in pathway flux to spare glucose for lactose synthesis, support energy generation from amino acids, and redirect amino acid carbon skeletons, oxaloacetate and acetyl CoA out of the mitochondria for use in triacylglycerol synthesis in the cytosol. Shifts in cytoskeletal proteins indicated a change to a polarized and stable phenotype and to support vesicular transport and transcriptional activity of the cell. Increased abundance of many transcriptional regulators and RNA processing proteins suggested an increased

output of transcripts, while the majority of changes involving protein processing implied the opposite – decreased protein biosynthesis and processing.

2.4.1 Metabolism

Changes in abundance of energy metabolism enzymes and proposed metabolic pathways in use in the differentiated state of the HC11 cells are shown in Figure 3 and should be referenced as the discussion proceeds. All differentially expressed proteins mentioned are also listed in Table 1 of the results section.

Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and phosphoglycerate kinase (PGK1) are both increased in differentiated cells, appearing to support the possible increase in the process of glycolysis as their reactions reversibly convert glyceraldehyde 3-phosphate to 1,3-bisphosphoglycerate, and 1,3-bisphosphoglycerate to 3-phosphoglycerate, respectively. The increase of pyruvate kinase (PKM) seen previously in multiple studies would substantiate the idea of increased glycolysis leading to high levels of pyruvate to be used for ATP production, all driven by a high glucose supply (Desrivieres et al., 2003; Rudolph et al., 2007; Strand, 2012). In this particular study, however, PKM was actually decreased in differentiated cells, leading to an alternative explanation of glucose sparing and energy production through other means. Down regulation of PKM would lead to a buildup of its substrate, phosphoenolpyruvate (PEP), as well as a decreased production of pyruvate from PEP. Phosphoglycerate kinase and GAPDH were both increased in the differentiated cells, but these enzymes can catalyze the conversion of their substrates and products reversibly. The accumulation of PEP, down regulation of PKM and increase of PGK1 and GAPDH would drive the direction of

glycolytic enzyme activity to move toward the product glyceraldehyde 3-phosphate (G3P). This pool of G3P could then serve several purposes. Some G3P could be converted to dihydroxyacetone phosphate (DHAP) and be used for triacylglycerol (TAG) synthesis. A portion of G3P could also be converted to glucose 6-phosphate (G6P) and shuttled into the pentose phosphate shunt (PPS) in order to produce the NADPH needed for fatty acid synthesis – a possibility supported by the up regulation of the PPS enzyme transketolase (TKT) in the differentiated state. This conversion to G6P would also allow more of the glucose in the cell to be directed toward lactose synthesis, which is in agreement with the previous findings that glucose phosphate isomerase has been shown to decrease in transcript abundance at onset of lactation, which would decrease the entrance of G6P into glycolysis (Rudolph et al., 2007).

If the enzyme abundance indicates the use of glucose for lactose synthesis and reverse glycolytic substrate conversion so as to shuttle substrates away from entering the TCA cycle, it is important to consider source of energy production and level of energy utilization in the MEC. Though the enzymes mentioned above have changed in abundance from the pregnant to the lactating state of the MEC, they are not nonexistent in one state or the other. A small amount of glucose is still being used for energy and a portion is still being transported into the PPS. Also, the energy production seems to be reduced from the proliferative pregnant state of the cell, which very well may require a higher level of ATP output relative to the differentiated lactating-like state. This idea is supported by the decrease in ATP synthase subunit alpha (ATP5A1), which is part of the F1, catalytic portion of the ATP synthase enzyme and required for the production of ATP in the respiratory chain (Wang and Oster, 1998).

In addition to requiring less energy in total, the differentiated cell appears to use more amino acids for energy and nutrient production. Serine hydroxymethyltransferase 2 (SHMT2) is increased in abundance and catalyzes the production of serine from glycine in the mitochondria, which can then be converted to 2-phosphoglycerate and transported into the cytosol where it can enter the glycolytic pathway to be used for production of energy or milk components. While production of pyruvate through means of glycolysis may be decreased due to down regulated PKM, it can also be made from threonine in a series of three conversions of which the last is catalyzed by aldehyde dehydrogenase (ALDH2); ALDH2 is increased in differentiated HC11 cells. Several acyl CoA dehydrogenases (ACAD) were also up regulated. These enzymes could be a part of fatty acid remodeling, or could serve to produce acetyl CoA from lysine in the mitochondria. Lysine is converted to α-ketoadipate in the cytosol and transported into the mitochondria, changed into glutaryl CoA and acted upon by ACAD, eventually leading to the production of acetyl CoA (Salway, 1994). This provides a pool of acetyl CoA in the mitochondria that is produced from a source other than pyruvate. Arginase (ARG1) abundance has repeatedly been shown to increase in the lactating mammary gland, including in this experiment (Yip and Knox, 1972; Oka and Perry, 1974; Mezl and Knox, 1977). The absence of other enzymes involved in the urea cycle has indicated an alternate role in the production of proline and glutamate from arginine. Other enzymes in these pathways have also been shown to increase in the lactating mammary gland and are thought to be necessary for production of milk proteins (Yip and Knox, 1972; Mezl and Knox, 1977).

Complementing the increased use of amino acids for energy and decreased overall need

for ATP, increased abundance of pyruvate carboxylase (PC), which converts pyruvate into oxaloacetate, and of aconitase (ACO2), which catalyzes the bidirectional conversion of citrate and isocitrate, supports increased production of citrate from oxaloacetate and acetyl CoA. In the rodent mammary gland, the citrate can be shuttled out of the mitochondria and the enzyme citrate lyase would catalyze the splitting of citrate back into oxaloacetate and acetyl CoA in the cytosol. This shuttling of citrate out of the mitochondria to allow acetyl CoA escape from the mitochondria for use in fatty acid synthesis has been documented, but is usually in concert with conversion of oxaloacetate into malate by cytosolic malate dehydrogenase (MDH1), and then into pyruvate for reentrance into the TCA cycle as part of the pyruvate/malate cycle, which is a means of NADPH production (Anderson et al., 2007; Rudolph et al., 2007). In this study MDH1 was down regulated, so the excess oxaloacetate produced must be directed in an additional, or different direction. Instead, the oxaloacetate could be converted into PEP by PEPCK in the cytosol and contribute to the reverse activity of glycolytic enzymes and continued increase of TAG synthesis, the PPS, and glucose sparing for lactose production.

Energy metabolism changes found in the transformation of HC11 cells from the undifferentiated to differentiated state largely support current knowledge of rodent mammary metabolism with just two enzyme regulation changes that seem to differ from previous findings, but may bring into question aspects of metabolic pathway flux. The decreased abundance of PKM and MDH1 in the differentiated state has not previously been detected and indicates a redirection of substrates in the glycolytic pathway and use of alternate materials for pyruvate production. This discovery may be due to differing

nutrient availability or differences in this HC11 cell line with other models and warrants further investigation.

2.4.2 Cytoskeleton

Changes in the profile of cytoskeletal proteins suggested increased stability of the cell and active vesicular trafficking as well as decreased need for filaments required for cell motility. Most of the keratins found, which included 5, 7, 14 and 17, were down regulated while the only up regulated keratin was 15. This was an interesting find that seems to be in agreement with some literature because keratin 15 (KRT15) has been found down regulated in proliferating cells and up regulated in differentiated, stable cells, while others such as keratins 5 and 14 specifically are associated with mitotically active cells and are down regulated in differentiated and phenotypically stable cell types (Porter et al., 2000; Waseem et al., 1999). Interestingly, although this has been found in certain studies in keratinocytes, a study on human mammary gland proposes KRT15 as a marker for a luminal progenitor cells and found it to be absent in differentiated lactating cells but present in undifferentiated cells (Moriera et al., 2010). This is an important difference to investigate and may be based on difference in species or cell source.

Prelamin-A (LMNA) was up regulated in the differentiated mammary epithelial cell. This protein can be in several forms as a precursor to mature lamin A and this precursor has been found to be responsible for several roles such as localization of other proteins to the nucleus and regulation of transcription due to transcription factor sequestration. It has been found to interact in this way with SREBP1, which is known to play a role in regulating lipid metabolism (Lattanzi, 2011).

Integrin beta-4 (ITGB4) serves with integrin alpha-6 as a receptor for laminins on the basement membrane and is thought to exist primarily on basal cells and to not be present on most luminal cells (Yang et al., 2009). In this study ITGB4 was down regulated in differentiated HC11 cells, which is interesting because the line arose from a luminal mammary epithelial cell and is seen in this study not only to express it, but also to change its expression with hormone treatment directing a differentiation-like process.

Gelsolin (GSN) was found to be both up and down regulated, meaning the predominating post-translational modified form of GSN was changed. Gelsolin is primarily known for its actin regulatory actions due to its ability to both sever and cap actin filaments to aid in cell mobility, but it also has a role in the progression of apoptosis. It has been shown to inhibit the release of cytochrome *c* from mitochondria, disallowing caspase-3 activation and inhibiting early stages of apoptotic signaling, therefore stabilizing the cell (Koya et al., 2000). Perhaps the up regulation of this form of GSN is occurring and allowing the differentiated cell to remain stable until involution. If the up regulated form of GSN is due to actin remodeling roles, perhaps the rearrangement of actin filaments is necessary to allow for vesicular trafficking and transport of milk components out of the cell. This is supported by the congruent up regulation of beta-actin and septin 8 in the differentiated HC11 cells. Septin 8 has been associated with vesicular trafficking and exocytosis in neural cells and is involved in secretory processes in the prostate, testis and ovary (Blaser et al., 2003; Ito et al., 2009; Shiryaev et al., 2012); septin 8 is possibly playing the same role in the mammary epithelial cell, aiding in transport and exocytosis of milk components.

2.4.3 Protein biosynthesis and processing

Of the twenty-two differentially expressed proteins involved in the biosynthesis and post-translational processing of protein, only four were up regulated in the differentiated cells. The majority of proteins was down regulated and involved the management of tRNAs, ribosome synthesis, and chaperoning other proteins. Of the four up regulated proteins, two of them were protein disulfide-isomerases (P4HB and PDIA4), which help rearrange disulfide bonds. This finding is interesting because the major whey protein in mouse milk is whey acidic protein (WAP). This protein is cysteine-rich and contains two separate domains, joined by disulfide bonds (Hennighausen and Sippel, 1982; Neville and Daniel, 1987). This increase could indicate a shift in protein product priority. The others were procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (PLOD2) and cathepsin D (CTSD) which form sites for carbohydrate attachment for collagen crosslink stability and break down proteins, respectively (Mercer et al., 2003; Benes et al., 2008). This mass down regulation of protein processing capability may be due to either a decreased overall amount of protein being produced in the differentiated state relative to the undifferentiated state, or a large, but more directed production of fewer specific proteins to be used in the production and secretion of milk components.

2.4.4 Transcriptional regulation and RNA processing

Transcript production and processing appeared to be increased immensely. There were a total of 9 proteins found to change in abundance; 3 proteins were down regulated and 6 proteins were up regulated. The up regulated proteins were mostly involved in processes indicative of a higher number of transcripts being present in the cell. Four different

proteins in the heterogeneous nuclear ribonucleoprotein (HNRNP) family were found to differ in abundance; A3, A2B1 and H2 were up regulated and AB was down regulated. HNRNPA2B1 and HNRNPH2 are involved in pre-mRNA processing (Dreyfuss et al., 2002; He and Smith, 2009) and A3 is involved in the cytoplasmic trafficking of RNA and mRNA maturation (Ma et al., 2002; Papadopoulou et al., 2012). The down regulated HNRNPAB acts as a repressor of transcription (Kamada and Miwa, 1992). TAR DNA-binding protein 43 (TARDBP) is up regulated and has been shown to interact with several hnRNP members and to assist in mRNA processing and stability (Buratte and Buralle, 2008). Max-like protein X (MLX) was also increased in abundance and is known to interact with another protein, MondoA. This heterodimer has been implicated in the transcriptional regulation of many metabolic genes, specifically glycolytic enzymes (Sans et al., 2006).

2.4.5 Comparison to Previous Studies

The changes in functional groups in HC11 seen in this study were largely in agreement with a much earlier study performed in the same cell line. Desrivieres and colleagues made the same comparison in 2003 using the HC11 cell line and even given the methods and time difference, a similar story was told. Protein folding and stability proteins were mostly down regulated while RNA processing and metabolic proteins were both largely up regulated in the differentiated HC11 cells. Cytoskeletal proteins were split in their regulation with many keratins being down regulated. There were many proteins identified in each experiment that were not identified in the other, but the general picture that each tells appears to agree and the majority of the proteins that were identified in both studies

shared direction of regulation. The level of similarity allowed confidence in the repeatability of 2DGE experiments and the integrity of the line, but there were a couple differences worth mentioning. The first small difference was that PDI was down regulated in Desrivieres' differentiated HC11 cells whereas it was up regulated in the present study. This does not have many clear implications, but was a specific difference in the two studies. In the Desrivieres study, PKM was up regulated in differentiated cells whereas it was down regulated in the present HC11 study. As previously mentioned, this is an interesting difference because its level of abundance potentially changes the flow of substrates through glycolysis and the direction of energy metabolism. This difference in regulation between these two studies of HC11 differentiation indicates that this regulation change does not likely reflect the cell source and most likely differs due to environment.

Proteome changes in HC11 MEC were very different in all major functional categories from those observed during differentiation in the MAC-T bovine MEC line (Strand, 2012). MAC-T showed opposite regulation of several specific protein processing and cytoskeletal proteins. Very few proteins involving RNA processing were even identified in the MAC-T cells so a proper comparison could not be made. Metabolic proteins differed quite a bit, but seemed to accurately demonstrate differences in the metabolic enzyme profiles of bovine and murine species. MAC-T cells indicated forward glycolytic enzyme activity, build-up of isocitrate and increased PPS activity. Data from a primary bovine comparison would be an important contribution to the comparison in order to accurately distinguish cell line and specie nuances.

HC11 cells and primary MEC isolated from pregnant and lactating mice (Strand, 2012) exhibited similar changes in proteins involved in metabolism and protein synthesis and

folding, but opposite regulation of RNA processing and cytoskeletal proteins. Both the primary cells and cell line displayed increases in glycolytic enzymes and citrate production due to up regulation of ACO2. Though not identified in the cell line, citrate lyase was found to be up regulated in the primary cells, supporting the idea of citrate shuttling out of the mitochondria and splitting into oxaloacetate and acetyl CoA for use in fatty acid and TAG synthesis, as previously proposed.

Interestingly, TKT was down regulated in the primary cells vs. the up regulation of TKT in the cell line, perhaps because the primary cells were obtaining some fatty acids from the adipocytes and the FA production rate was not as high as in the cell line, requiring less NADPH. Pyruvate dehydrogenase was up regulated and PKM was not identified in the primary cells, suggesting a possible flow of substrates down the glycolytic pathway and into the mitochondria and TCA cycle. Respiratory chain energy production also seemed to be increased due to the increase in ATP5A1 in primary lactating cells. Though the regulation of many enzymes involved in energy metabolism are left unknown through proteomics and absence of their finding does not mean they did not change, speculations can be made based on findings that did occur. Regulation of enzymes found in the primary cells seem to suggest a possible increased use of glucose for glycolysis and energy production relative to the cell line. This possible difference in metabolism could be explained in several ways. Signals from adjacent cells or even hormone signals that were present in the mammary gland of the mice before MEC isolation could have influenced regulation of different metabolic enzymes in the primary cells. The primary cells could have also been provided higher levels glucose prior to isolation through diet, glycogenolysis and/or gluconeogenesis in the liver of the mouse and this excess glucose

was then available for use in lactose synthesis, energy production, and other milk component production.

Speculation about the difference in the cytoskeletal regulation and transcriptional regulation is more difficult. More specific cytoskeletal component comparisons would be useful as there were only two shared between the two models, LMNA and KRT5, and these were oppositely regulated. The environment in which each model was maintained was very different and surely influences the cytoskeletal components needed to maintain adhesion and structure of the cells, suggesting explanation for differences occurring. The RNA processing proteins were oppositely regulated which could also be a consequence of their environments. Perhaps the cell line has increased production of more transcripts due to the lack of negative regulation from surrounding cell types, or they may actually require the extra transcripts due to the lack of paracrine factors or other needs due to environment.

This comparison shows that HC11 cells, which are currently widely used to investigate events occurring in the MEC, may need to be analyzed more closely before trusting accuracy of findings. This study showed a striking difference in RNA processing and cytoskeletal dynamics between this cell line and primary MEC and some subtle yet meaningful differences in energy metabolism, which is reason to question the accuracy of the HC11 cell line to use as a model for certain aspects of MEC differentiation.

CHAPTER 3 – Proteomic analysis of isolated primary virgin, pregnant and primiparous quiescent mammary epithelial cells

3.1 Introduction

Investigation of the molecular events occurring throughout mammary development is an important and heavily studied topic because of its potential influence on the dairy industry, breast feeding, and the fight against breast cancer. Technologies aimed at obtaining large and undefined data sets such as microarrays and 2DGE have been used to analyze the changes occurring in the gland, which provide the ability to discover new molecules and processes involved that were previously not connected to mammary development. Most of these studies perform analysis on RNA or protein extracts obtained from whole tissue, which includes all information from not only the parenchyma but also the adipose, immune cells, endothelium and extracellular matrix (Rudolph et al., 2003; D'cruz et al., 2002; Beddek et al., 2003). This approach provides insight into changes in the entire gland, but does not give a clear picture of events occurring in the mammary epithelial cells and requirements for their differentiation into lactating cells. Other –omic investigations have analyzed changes during the differentiation of cell lines, such as the murine HC11 and the bovine MAC-T lines (Desrivieres et al., 2003; Strand, 2012). Data provided by these studies is more representative of the MEC itself, but there are also limitations of analyzing cell lines and *in vitro* experimental data. Cell lines have potential of changing over time and expressing transcripts and proteins unrepresentative of primary MEC and any data obtained *in vitro* is generated lacking influence from surrounding tissues and the natural environment for the MEC. Analysis of primary MEC from different stages of mammary development would eliminate the limitations of in vitro and

whole gland investigations by being representative of the MEC and receiving cues from their true environment throughout development.

Previous investigation into changes occurring between pregnant and lactating states of the primary murine MEC has been performed in the Peterson lab (Strand, 2012). Results of this previous investigation were compared with the *in vitro* induction of a differentiated-like state of the HC11 murine MEC line by hormone treatment to elucidate similarities and differences between these two models of MEC differentiation; this comparison was discussed in the previous chapter. Changes in the mammary gland and MEC occur between all stages of development including from the virgin state to the pregnant state and from the lactating state back to quiescence. The following investigation aims to add to the data set from primary isolated MEC in different stages of mammary development. Primary MEC from the virgin state are isolated and compared to MEC from the pregnant state and primiparous quiescent state of the gland. The discovery of protein changes will add to the knowledge of the events that occur in preparation for the shift to a lactating phenotype and unveil changes in protein expression that persist in the MEC after full development, function, and involution of the mammary gland.

3.2 Methods

3.2.1 Animals and breeding

All procedures were approved by the Cal Poly Institutional Animal Care and Use Committee. ICR mice (Taconic, Hudson, NY) were housed with a 12 h light schedule and *ad libitum* access to food and water. Samples were taken from virgin and pregnant mice between 10 and 11 weeks of age. Pregnant samples were taken at day 10 of pregnancy. Due to resource and scheduling constraints, primiparous samples were taken from mice approximately 23 weeks in age. Primiparous mice were weaned at 21 days of lactation and samples were taken at 18 days post-weaning. Estrous cycle was not taken into account for virgin or primiparous samples. All mice were euthanized by CO_2 asphyxiation followed by cervical dislocation. Mammary tissue was harvested from all glands. Samples consisted of all glands pooled from one mouse. Three samples for each stage of development were used.

3.2.2 Primary mammary epithelial cell isolation

Immediately following euthanasia all mammary tissue was removed and rinsed in 1X Hank's Balanced Salt Solution. Tissue was then transferred and minced in digestion media containing collagenase, trypsin and EDTA in Dulbecco's Modified Eagle Medium. Media containing tissue was placed in an incubated shaker set at 37°C for 90 minutes, with disruption by pipet every 30 minutes. Cells were pelleted and red blood cells were removed by incubation with blood cell lysis buffer (8.3 g/L ammonium chloride in 0.01M Tris-HCl). The remaining pellet was then plated in T-75 flasks for one hour in order to remove the rapidly adherent fibroblasts. The cell suspension was then incubated with EDTA and DNase solutions before being filtered through a 100um filter. Epithelial cells were re-suspended in DMSO cell freezing media, brought down to -80°C and stored in a liquid nitrogen tank until extraction.

3.2.3 Protein extraction

Protein extractions were performed the same as was specified in section 2.2.2.

3.2.4 Two-dimensional gel electrophoresis (2DGE)

Two-dimensional gel electrophoresis was performed the same was as specified in section 2.2.3.

3.2.5 Gel analysis, spot picking and trypsin digestion

Gel analysis, spot picking and trypsin digestion were performed the same as specified in section 2.2.4.

3.2.6 Mass spectrometry and protein identification

Mass spectrometry and protein identification were performed the same as specified in section 2.2.5.

3.3 Results

All gels containing virgin, pregnant, and primiparous quiescent primary cell protein extracts were scanned and overlaid, creating one fused image for use in spot detection. Delta2D analysis of the fused gel image detected a total of 234 individual protein spots (Figure 4). Comparisons were then made between the virgin and pregnant gels and between the virgin and primiparous quiescent gels. Of the detected spots, 67 and 73 of these spots differed in abundance between the virgin and pregnant sample comparison and the virgin and quiescent sample comparison, respectively ($P < 0.02$). Of these differentially expressed spots, 31 and 36 were identified for the virgin and pregnant comparison (Table 2) and the virgin and quiescent comparison (Table 3), respectively when using the MASCOT server and matching the spectra to the NCBI Database. Of the 31 identified proteins in the virgin and pregnant comparison, 3 were up and 28 were down regulated in the pregnant MEC relative to the virgin MEC. Of the 36 identified proteins in the virgin and pregnant comparison, 5 were up regulated and 31 were down regulated in the primiparous quiescent MEC relative to the virgin MEC.

Functional groups for the majority of the identified differentially expressed proteins in the virgin and pregnant MEC comparison include energy metabolism (11 spots total; 1 up regulated/10 down regulated), cytoskeletal dynamics (7 spots total; 1 up regulated/6 down regulated), and protein synthesis and processing (6 spots total; 1 up regulated/5 down regulated).

Functional groups for the majority of the identified differentially expressed proteins in the virgin and primiparous quiescent MEC comparison include energy metabolism (9 spots total; 1 up regulated/ 8 down regulated), cytoskeletal dynamics (7 spots total; 3 up

regulated/4 down regulated), protein synthesis and processing (9 spots total; 1 up regulated/8 down regulated), and transcriptional regulation and RNA processing (3 spots total; all down regulated).

Figure 4. *Fused gel image of isolated primary mammary epithelial cell protein spots*. Protein was extracted from isolated primary virgin, pregnant, and primiparous quiescent mouse mammary epithelial cells. Proteins were subjected to 2-dimensional gel electrophoresis, stained with Coomassie blue, and analyzed by Delta 2D. The image represents all gels fused into one image. Labeled spots were differentially expressed between treatments ($P < 0.02$). Numbers correspond to "Spot ID" in Table 2 and Table 3.

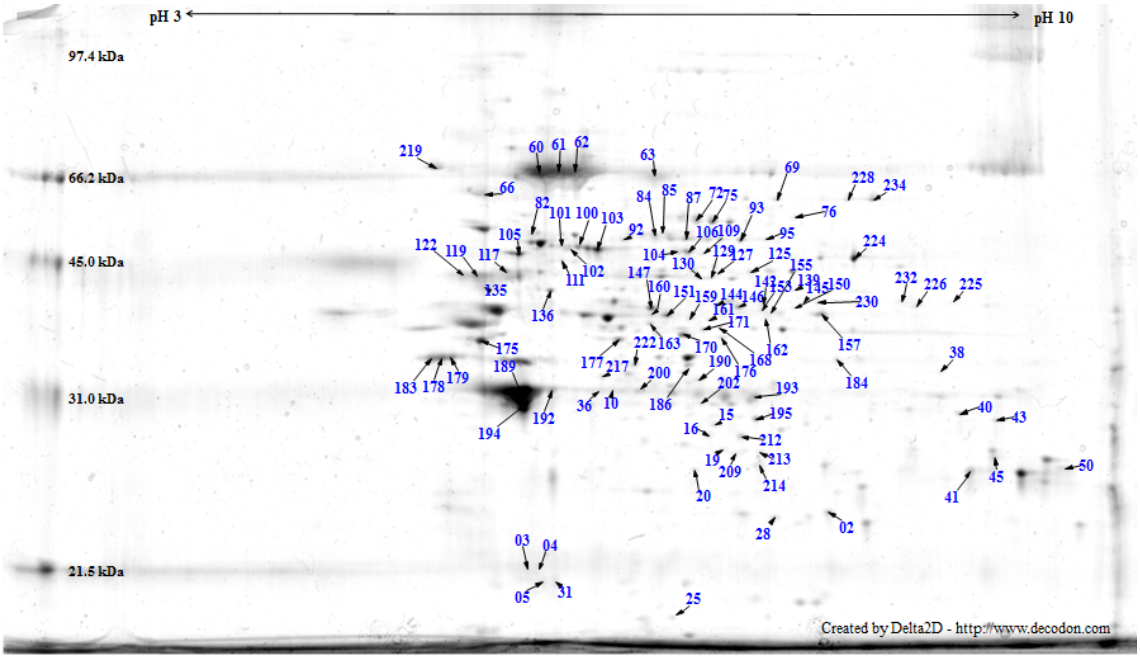

Figure 5. *Heat map of protein spots that were differentially expressed between isolated primary MEC from virgin and pregnant mice*. Created by hierarchical clustering using Pearson's correlation of HC11 protein spots that were differentially expressed between the undifferentiated and differentiated states ($P < 0.02$). Blue represents lower than average protein abundance, while orange represents higher than average protein abundance for that particular spot. Rows represent individual spots and the numbers correspond to "Spot ID" in Table 2. Columns represent individual gels. Virgin treatment gels are on the left and pregnant treatment gels are on the right.

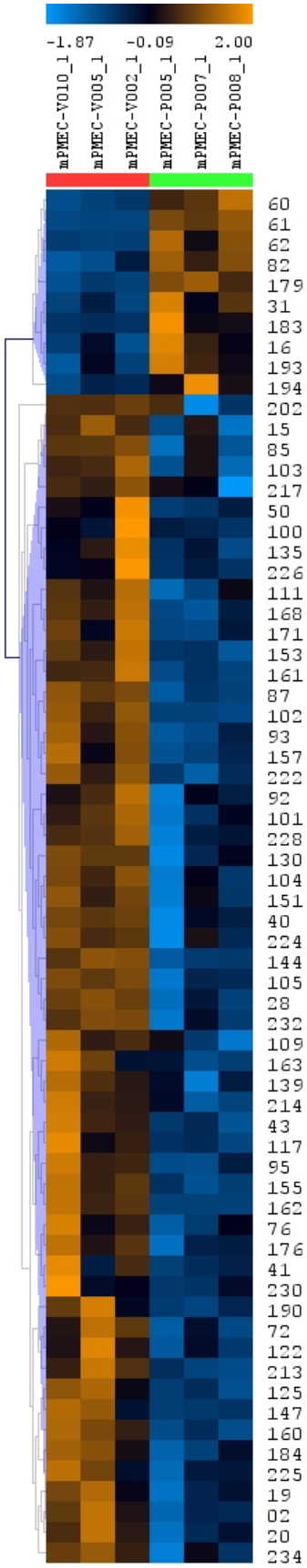

Figure 6. *Heat map of protein spots that were differentially expressed between isolated primary MEC from virgin and primiparous quiescent mice*. Created by hierarchical clustering using Pearson's correlation of HC11 protein spots that were differentially expressed between the undifferentiated and differentiated states ($P < 0.02$). Blue represents lower than average protein abundance, while orange represents higher than average protein abundance for that particular spot. Rows represent individual spots and the numbers correspond to "Spot ID" in Table 3. Columns represent individual gels. Virgin treatment gels are on the left and primiparous quiescent treatment gels are on the right.

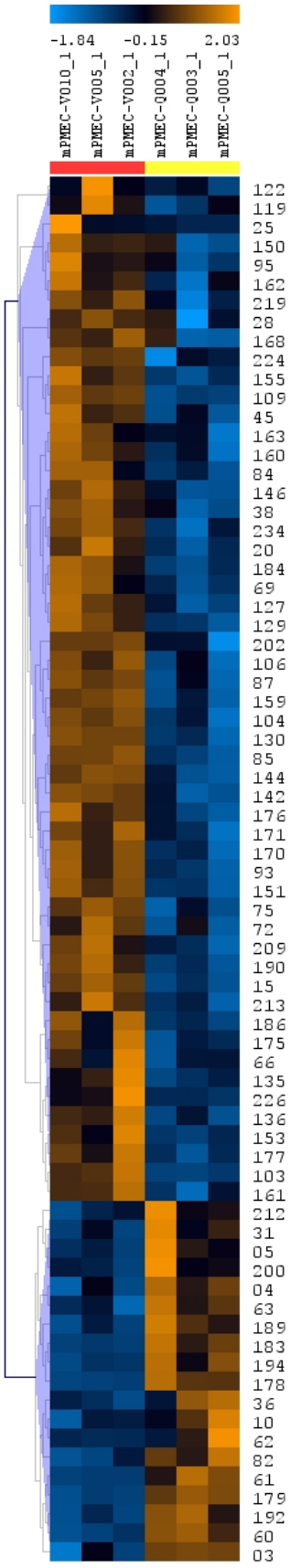

Table 2. *List of identified protein spots that were differentially expressed between isolated primary virgin and pregnant MECs* ($P < 0.02$). Under "Direction of Regulation," "up" indicates an increased abundance of the specified protein, while "down" means the decreased abundance of the specified protein in the pregnant cells relative to the virgin cells. "Fold change" indicates magnitude of abundance change and was calculated by dividing mean spot abundance of pregnant gels by mean spot abundance of virgin gels.

Spot ID	Protein Name	Gene Name	Direction of Regulation	Fold change
	cytoskeleton			
95	keratin, type II cytoskeletal 1	Krt1	down	0.35
151	keratin, type I cytoskeletal 16	Krt16	down	0.59
194	keratin, type I cytoskeletal 19	Krt19	up	1.39
19	LIM and SH3 domain protein 1	Lasp1	down	0.63
93	prelamin-A/C isoform A precursor	Lmna	down	0.49
105	lamin-B1	Lmnb1	down	0.69
20	PDZ and LIM domain protein 1	Pdlim1	down	0.56
	immune function			
230	catalase	Cat	down	0.47
202	immunoglobulin heavy chain variable region	Ighvdj	down	0.73
	ion transport			
2	voltage-dependent anion channel 2	Vdac2	down	0.64
	metabolism			
193	long-chain specific acyl-CoA dehydrogenase, mitochondrial precursor	Acadl	up	1.65
15	short-chain specific acyl-CoA dehydrogenase	Acads	down	0.75
40	acetyl-CoA acetyltransferase, mitochondrial precursor	Acat1	down	0.71
228	aconitate hydratase, mitochondrial precursor	Aco2	down	0.56
234	aconitate hydratase, mitochondrial precursor	Aco2	down	0.66
43	fructose-bisphosphate aldolase A isoform 1 precursor	Aldoa	down	0.68
28	electron transfer flavoprotein subunit alpha, mitochondrial	Etfa	down	0.46
41	glyceraldehyde-3-phosphate dehydrogenase	Gapdh	down	0.64
50	malate dehydrogenase	Mdh2	down	0.55
224	transketolase	Tkt	down	0.79
92	heat shock protein 75 kDa, mitochondrial	Trap1	down	0.56
	plasma protein			
103	alpha-fetoprotein, partial	Afp	down	0.58
100	serum albumin	Alb	down	0.53

Spot ID	Protein Name	Gene Name	Direction of Regulation	Fold change
160	serum albumin	Alb	down	0.52
72	serotransferrin precursor	Tf	down	0.46
	protein synthesis & processing			
102	heat shock protein 1A	Hspa1a	down	0.40
82	heat shock protein 70 cognate	Hspa8	up	3.74
117	alpha-1-antiproteinase precursor	SERPINA1	down	0.53
122	alpha-1-antiproteinase precursor	SERPINA1	down	0.50
135	alpha-1-antiproteinase precursor	SERPINA1	down	0.41
190	elongation factor Tu, mitochondrial isoform 2	Tufm	down	0.72

Table 3. *List of identified protein spots that were differentially expressed between isolated primary virgin and primiparous quiescent MECs* ($P < 0.02$). Under "Direction of Regulation," "up" indicates an increased abundance of the specified protein, while "down" means the decreased abundance of the specified protein in the primiparous cells relative to the virgin cells. "Fold change" indicates magnitude of abundance change and was calculated by dividing mean spot abundance of primiparous quiescent gels by mean spot abundance of virgin gels.

Spot ID	Protein Name	Gene Name	Direction of Regulation	Fold change
	cytoskeleton			
189	actin, cytoplasmic 1	Actb	up	1.47
192	actin, cytoplasmic 1	Actb	up	2.34
95	keratin, type II cytoskeletal 1	Krt1	down	0.59
151	keratin, type I cytoskeletal 16	Krt16	down	0.51
194	keratin, type I cytoskeletal 19	Krt19	up	1.87
93	prelamin-A/C isoform A precursor	Lmna	down	0.43
20	PDZ and LIM domain protein 1	Pdlim1	down	0.55
	DNA replication & cell division			
25	cell division control protein 42 homolog	CDC42	down	0.36
	immune function			
202	immunoglobulin heavy chain variable region	Ighvdj	down	0.60
	ion transport			
2	voltage-dependent anion channel 2	Vdac2	down	0.68
	metabolism			
15	short-chain specific acyl-CoA dehydrogenase	Acads	down	0.63
234	aconitate hydratase, mitochondrial precursor	Aco2	down	0.61
170	aldehyde dehydrogenase, mitochondrial precursor	Aldh2	down	0.48
175	ATP synthase subunit beta, mitochondrial precursor	Atp5b	down	0.55
146	dihydrolipoamide dehydrogenase precursor	DLD	down	0.46
28	electron transfer flavoprotein subunit alpha, mitochondrial	Etfa	up	1.12
38	fumarate hydratase, mitochondrial precursor	Fh	down	0.81
106	succinate dehydrogenase flavoprotein subunit	Sdha	down	0.34
224	transketolase	Tkt	down	0.66
	plasma protein			
103	alpha-fetoprotein, partial	Afp	down	0.60
160	serum albumin	Alb	down	0.60
72	serotransferrin precursor	Tf	down	0.47
75	serotransferrin precursor	Tf	down	0.50
84	serotransferrin	Tf	down	0.44
	protein synthesis & processing			
186	elongation factor 1-gamma	Eef1g	down	0.42
66	heat-shock protein 84	Hsp90ab1	down	0.48

Spot ID	Protein Name	Gene Name	Direction of Regulation	Fold change
219	endoplasmin	Hsp90b1	down	0.42
82	heat shock protein 70 cognate	Hspa8	up	3.93
136	60 kDa heat shock protein, mitochondrial	Hspd1	down	0.60
122	alpha-1-antiproteinase precursor	SERPINA1	down	0.72
135	alpha-1-antiproteinase precursor	SERPINA1	down	0.33
119	alpha-1-antiproteinase precursor	SERPINA1	down	0.40
190	elongation factor Tu, mitochondrial isoform 2	Tufm	down	0.45
	transcriptional regulation & RNA processing			
45	heterogeneous nuclear ribonucleoprotein A3	Hnrnpa3	down	0.55
177	heterogeneous nuclear ribonucleoprotein H	Hnrnph1	down	0.57
69	far upstream element-binding protein 2	Khsrp	down	0.65

3.4 Discussion

In this study, the protein expression of primary isolated MECs from virgin mice was compared with that of primary isolated MECs from pregnant and primiparous quiescent mice of the same lineage. As previously mentioned, the methods of analysis used in 2DGE allow the discovery of protein changes occurring at a point for each stage of development investigated, but only provide a fraction of all changes occurring. The results for this set of experiments included an even smaller fraction that expected. The small number of identifications could have been a result of low protein abundance in the samples run, as original protein concentration was lower than desired and the procedures had to be repeated several times due to technical difficulties. Another possibility, though not as likely, is that the proteins have not yet been identified and did not match any NCBI database information. This low number of protein identities must be kept in mind when reading the following discussion, as it is more difficult to put together a possible scenario occurring in the cell with such a small subset of information. This being said, the trend of down regulation in both comparisons relative to the virgin MECs is likely a true occurrence. In the virgin to pregnant comparison, only 10 of the 67 differentially expressed spots were up regulated in the pregnant MEC. In the virgin to primiparous quiescent comparison, only 19 of the 73 differentially expressed spots were up regulated in the primiparous MEC. The spots that were successfully identified closely mirrored this trend of down regulation from the virgin state with only 5 out of 36 and 3 out of 31 spots being up regulated in the pregnant and primparous MEC, respectively.

In addition to the small data set, another variable to consider when reviewing the following information is the possibility of contaminating cell types. Though the aim of

the isolation procedures was to end up with isolated luminal mammary epithelial cells, there are many more cell types in the gland that could have remained in the final pellet. The most probable contaminating cell type would be myoepithelial cells, followed by endothelial cells and leukocytes. While the presence of adipocytes, fibroblasts, and erythrocytes is also possible, specific and effective steps were employed for their elimination and they possess phenotypes distinct from the MECs and were not seen when the cells were isolated and examined microscopically.

One last consideration is the stage of the estrous cycle in which the virgin and primiparous quiescent mice were harvested. This factor was not monitored or confirmed on the date of harvest, resulting in no knowledge regarding in which stage samples were taken or consistency in stage of cycle across these different samples. As mentioned in chapter 1, the activity of the cells does change with stage in estrous cycle and this would very likely have an effect on the protein expression profile of these cells.

Although there were disappointing limitations based on the data obtained, and the variables mentioned above must be considered when analyzing the resulting data, proposed scenarios based on the changes in the protein profiles in the different stages of development were very interesting. The seemingly high abundance of many protein types in the virgin MEC compared to both other stages of development were surprising and further investigation into this phenomenon would be a fascinating comparator. A discussion of both comparisons performed and the possible events occurring in these cells will follow.

3.4.1 Virgin vs Pregnant Comparison

3.4.1.1 Cytoskeleton

All but one of the identified proteins related to cytoskeletal dynamics were down regulated from the virgin to the pregnant state. This may first seem surprising as it may seem that differentiating cells would require reorganization and increased cytoskeletal function would be required, but when the individual proteins are investigated a possible logical scenario is revealed. PDZ and LIM domain protein 1 (PDLIM1, also known as CLIM1) and LIM and SH3 domain protein 1 (LASP1) were both down regulated from the virgin to the pregnant state of the MEC. PDLIM1 has been found in close association with hormone receptor, ERα and is a co-regulator of ERα's transcriptional targets. Its expression has been found very high in less-differentiated and aggressive breast tumors, and has also been found to block differentiation in several cell types, including MEC *in vitro* (Johnsen et al., 2009). LASP1 is an actin binding protein that is required for cell migration and is also highly expressed in breast tumors (Lin et al., 2004). The down regulation of these proteins in the primary pregnant MEC might indicate that the proliferation and reorganization of early pregnancy has ended and the differentiation of these cells into the lactating phenotype can be performed. Two lamin proteins were also down regulated in the pregnant state: prelamin A/C (LMNA) and lamin B1 (LMNB1). The lamin proteins have long been known as the major component of the nuclear lamina, providing structure and stability to the nuclear envelope as well as serve as anchor points for chromatin. Now they are thought to play other roles in processes such as DNA synthesis, transcription, and apoptosis (Goldman et al., 2002). The down regulation of these proteins is interesting, especially since the loss of LMNB1 has been associated with

cell senescence in previous investigations (Sadaie et al., 2013; Freund et al., 2012). Three keratins (KRT) were also identified. KRT1 and KRT16 were down regulated in the pregnant state, while KRT19 was one of the few proteins that were up regulated. KRT16 has been reported as a normal resident of the mammary gland (Pellegrino et al., 1988), while it was surprising to see differential expression of KRT1, as previous investigation has reported it as being absent in all stages of mammary development (Mikaelian et al., 2006). The up regulation of KRT19 was of interest due to the few proteins that showed this direction of expression. The increase in KRT19 in the pregnant state is in agreement with previous reports that its expression is highest during pregnancy and lactation. It is indicative of low proliferation and increased ability for milk component secretion, suggesting gland is in the differentiating stage following the spike in expansion mentioned in chapter 1 (Bartek et al., 1990).

Overall, the differential expression of the cytoskeletal proteins from virgin to the pregnant state imply that the cells isolated from the pregnant glands are in the differentiating stage following the highly proliferative stage of early pregnancy with decreased motility and increased indicators of differentiation.

3.4.1.2 Metabolism

All proteins involved in metabolic processes were down regulated in the pregnant state compared with the virgin state with the exception of one. Long-chain specific acyl-CoA dehydrogenase (ACADL) was increased in expression in the pregnant primary MEC, while short chain acyl-CoA dehydrogenase was decreased. Though many proteins in metabolic processes were not identified and their direction of regulation is unaccounted

for, the proteins that were identified suggested mass decrease in energy metabolism compared to the virgin state. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and fructose-bisphosphate aldolase A (ALDOA) down regulation suggest decreased glycolysis. Whether there is a lower influx, a build up and storage of glucose, or use for other processes at this point in development cannot be determined by the data gathered, but it does not appear that glucose is entering glycolysis at as high of a rate as in the virgin MEC. The trend of lower energy generation continues with a decrease seen in dihydrolipoamide dehydrogenase, aconitate hydratase (ACO2), and malate dehydrogenase, which function in the conversion of pyruvate to acetyl CoA and the TCA cycle. Malate dehydrogenase can also aid in the generation of NADPH, which also may be decreased based on the down regulation of transketolase, a member of the pentose phosphate shunt. The trend suggests that less energy generation is occurring in the pregnant state than the virgin state, possibly suggesting that differentiated state of the cell is more energy efficient. One last protein is a member of the heat shock proteins, but has been shown to have regulatory effects on metabolism. Heat shock protein 75 kDa (TRAP1) is down regulated in the pregnant state relative to the virgin state. This is interesting because of a previous report that deficiency in TRAP1 leads to an increase in mitochondrial respiration, fatty acid oxidation, and a buildup of TCA cycle intermediates, ATP and reactive oxygen species (ROS), while also suppressing glucose metabolism. In addition to metabolic effects, its decrease is also associated with increased cell invasiveness (Yoshida et al., 2013). While there is no evidence of increased respiration and fatty acid oxidation in our data, there is also no evidence of its suppression and there is evidence of suppression of glycolysis. The decrease in TRAP1 may be partially

responsible for depression of glucose metabolism in this investigation, though the cytoskeletal evidence suggests increased stability and polarization rather than mobility and invasiveness.

Overall the interesting trend of metabolic processes was a decrease in activity in the pregnant state of the cell, which means it was high in the virgin state. This is an interesting phenomenon that might be explained in several ways. Perhaps the virgin MEC requires more energy use due to the fluctuations in proliferation and regression seen with the estrous cycle, or maybe the difference in available nutrients in the blood stream due to the hormones of pregnancy has an effect on metabolic regulation and the use of glucose in the two states of the cell.

3.4.1.3 Protein synthesis and processing

In addition to the TRAP1, two other heat shock proteins were identified as differentially expressed in this comparison. Heath shock protein 1A (HSPA1A) is down regulated in the pregnant MEC. This is a stress-induced HSP and helps the cell manage aggregation of denatured proteins in stress conditions. In contrast, heat shock protein 70 cognate (HSPA8) was up regulated in the pregnant MEC. This protein is 86% identical to HSPA1A, but is a constitutively expressed protein that acts as a housekeeper for the cell, aiding in protein folding and transport as well as managing denatured protein aggregation in times other than just high stress (Daugaard et al., 2007). Perhaps the up regulation of HSPA8 balances the decreased need for HSPA1A and the state of the pregnant MEC is in a less stressful state than the virgin MEC. Though not many proteins in this category were identified, the ones that were indicate a lower amount of protein relative to the

virgin state. Alpha-1-antiproteinase (SERPINA1) and elongation factor tu (TUFM) are both expressed at a lower level in the pregnant state than the virgin state. These proteins inhibit serine proteases and aid in the production of proteins by helping with the binding of tRNA at the ribosome, respectively (Kalsheker, 1988; Stark et al., 1997). This would allow increased protein degradation and decreased protein production, though these are only two of a multitude of proteins involved in these processes for which we do not know the direction of regulation.

3.4.1.4 Secreted proteins

Interestingly, there were several identified proteins that are secreted and typically found circulating in the plasma that were identified in this comparison. Albumin (ALB), alpha-fetoprotein (AFP), and serotransferrin (TF) were all identified as down regulated in the pregnant state. The function of AFP is not well understood, though it is thought to be responsible for the transport of various ligands such as fatty acids and heavy metals. Roles as a growth promoter and a growth suppressor have both been assigned to this protein and it has been recognized as a protein present during pregnancy that has lasting protective effects from breast cancer risks (Vakharia and Mizejewski, 2000; Parikh et al., 2005). This is very interesting since this protein was decreased in the pregnant state versus the virgin state of the MEC. The protein responsible for iron transport, TF, was also decreased in the pregnant state. This protein is typically found in milk product during lactation, but the decrease in pregnancy from the virgin state in this investigation might be explained by the decreased need of iron at that moment of cell isolation. Iron is required as a cofactor during DNA replication and growth, which, as previously

mentioned, might not be occurring in the pregnant MEC at their stage of development (Gomme and McCann, 2005). Albumin is also a component of the final milk product and though its down regulation from the virgin state might be surprising, the lack of up regulation is not as the inclusion of protein transport into the alveolar lumen may not yet have began.

3.4.1.5 Other identified proteins

Voltage-dependent anion channel 2 (VDAC2) was found decreased in abundance in the pregnant state relative to the virgin state of the MEC. This protein functions as an ion channel in the mitochondrial outer membrane and is responsible for transport of metabolites such as ATP and ADP across the membrane (Rostovtseva and Bezrukov, 2008). Perhaps since the metabolism appears suppressed in the pregnant cells, the regulation and transport of metabolites is also reduced and VDAC2 is decreased due to lack of necessity. Another interesting expression profile of the pregnant state relative to the virgin state is the decrease in abundance of two proteins related to immune and stress-fighting function. Catalase and immunoglobulin heavy chain, variable region (IGHVDJ) were both found down regulated in the pregnant MEC. Catalase is responsible for protection against reactive oxygen species and IGHVDJ is the heavy chain portion of immunoglobulin. The decrease in catalase agrees with the story told by the heat shock proteins, that possibly the pregnant MEC does not require as much of this protein due to a less stressful state. It is interesting that IGHVDJ is not only present in the virgin and pregnant cells, but that it varies in regulation as this protein is only produced by B cells of the immune system. Immunoglobulins are transported from the blood into the final milk

product, but this was thought to occur near parturition. The explanation of a contaminating cell type during the virgin isolation should not be overlooked in this situation.

3.4.2 Virgin vs Primiparous Quiescent Comparison

3.4.2.1 Cytoskeleton

The identified cytoskeletal protein expression profile for the primiparous quiescent MEC varies only slightly from the pregnant MEC relative to the virgin state of the cells. LASP1 and LMNB1 were the only two present in the virgin/pregnant comparison that were not identified as differentially expressed between the primiparous quiescent and virgin states. PDLIM1, LMNA, and the keratins (1, 16, 19) all show the same direction of regulation as seen in the virgin/pregnant comparison above. This similar expression pattern for the pregnant and quiescent states relative to the virgin state appear to represent persisting differences in the MEC after one pregnancy and lactation event. Since PDLIM1 has been implicated in aggressiveness of breast cancer (Johnsen et al., 2009), this continued decrease in abundance could support the idea that it may play a role in tumor metastases and decreased risk following a pregnancy. In the same way, the increased expression of KRT19 indicates a continued state of a higher order of differentiation, which has been seen histologically and is a well-documented phenomenon (Bartek et al., 1990; Richert et al., 2000; Daniel and Silberstein, 1987). In addition to the proteins that were identified in the previous comparison, actin, cytoplasmic 1 (β β -actin; ACTB) also appeared as up regulated in the quiescent state

relative to the virgin state. Increased actin is surprising as it indicates cell movement or cytoskeletal rearrangement. Perhaps as beta-actin was increased in the differentiated HC11 cells, the quiescent cells have persisting cytoskeletal changes from the previously lactating phenotype in which the cytoskeletal rearrangement was necessary for transport of milk components to the alveolar lumen.

3.4.2.2 Metabolism

The metabolic state of the primiparous quiescent MEC appears to be suppressed relative to the virgin state of the cells. The glycolytic enzymes identified in the last comparison were not indicated as differentially expressed between the virgin and primiparous quiescent states, so the regulation of this pathway in this investigation cannot be speculated upon. Enzymes involved in the TCA cycle and respiratory chain, however, were identified. Dihydrolipoamide dehydrogenase (DLD) is on of three enzymes that makes up the pyruvate dehydrogenase complex which converts pyruvate to acetyl CoA for entrance into the TCA cycle (Patel and Korotchkina, 2006). This means that there is either a lower amount of acetyl CoA entering the TCA cycle or the substrate is coming from another origin. DLD also plays a role in the conversion of 2-oxoglutarate to succinyl-CoA in the TCA cycle. Aconitate hydratase (ACO2), succinate dehydrogenase flavoprotein subunit (SDHA), and fumarate hydratase (FH) are also all down regulated relative to the virgin state. These all catalyze reactions of the TCA cycle. SDHA links the TCA cycle to the respiratory chain. In addition to the TCA cycle link being decreased, ATP synthase subunit beta is down regulated in the primiparous state. This is a part of the F1, catalytic portion of the ATP synthase enzyme and required for the production of ATP

in the respiratory chain (Wang and Oster, 1998). Interestingly, the only identified protein that was differentially expressed in both comparisons, but displayed opposite directions of regulation was electron transfer flavoprotein subunit alpha (ETFA). Increased abundance of ETFA in the primiparous state relative to the virgin state was detected in this investigation, while it was down regulated in the pregnant state relative to the virgin state. The role of ETFA is to transfer electrons from several dehydrogenases to the respiratory chain (Toogood et al., 2007). The increase in this protein and decrease in ATP5B suggests an interesting contrast in respiratory activity. Production of NADPH might also be reduced in the primiparous state as one of the major PPS enzymes, TKT, is down regulated. Perhaps the lipid metabolism pathways are suppressed altogether, as the ACADS is also reduced in this state. Aldehyde dehydrogenase (ALDH2) is the last identified metabolic protein that is down regulated in the primiparous state. Though mostly known for alcohol metabolism, ALDH2 can also participate in other metabolic processes such as the series of reactions resulting in the conversion of threonine to pyruvate (Salway, 1994), which was a proposed reason for its up regulation in differentiated HC11 cells in the previous chapter.

Overall metabolic activity in the primiparous quiescent MEC relative to the virgin MEC reveals a suppression of activity, much like what was seen in the virgin/pregnant comparison. This with the similarities in cytoskeletal proteins, which indicated a higher order of differentiation may suggest the greater requirement for energy by the less differentiated, virgin state of the cell.

3.4.2.3 Protein synthesis and processing

Similar to the previous comparison, HSPA8 was up regulated from the virgin state to the primiparous state. Heat shock protein 84 (HSP90AB1), endoplasmin (HSP90B1), and 60 kDa heat shock protein (HSPD1) were all down regulated in the primiparous state. HSP90B1 and HSPD1 function in the endoplasmic reticulum and the mitochondria, respectively, to participate in protein folding and transport (Chen et al. 2005; Koll et al., 1992). The cytosolic HSP90AB1 is constitutively expressed and involved in folding of cell regulatory proteins and refolding of stress-denatured proteins (Chen et al., 2005). Perhaps the up regulation of HSPA8 was enough to regulate stress-denatured proteins and the need for folding of cell regulatory proteins was decreased in the non-proliferative state of the primiparous relative to the virgin MEC. Elongation factor 1-gamma (EEF1G) and TUFM were also both down regulated in the primiparous MEC. This indicates decreased translation as these elongation factors aid in the delivery of tRNAs to the ribosome to allow production of proteins. Lastly, SERPINA1 was again found down regulated relative to the virgin state of the MEC. Just as in the pregnant MEC, the regulation of the identified proteins related to protein processing occurring in the primiparous MEC suggested a less stressful environment with decreased protein production and the potential for increased degradation.

3.4.2.4 Transcriptional regulation and RNA processing

In addition to the decrease in protein abundance, three proteins involved in transcriptional regulatory processes were also down regulated. Two heterogeneous nuclear ribonucleoprotein members, A3 and H, were decreased in abundance in the primiparous

state relative to the virgin state. These proteins are involved in the cytoplasmic trafficking of RNA, pre-RNA processing, and mRNA maturation (Ma et al., 2002; Papadopoulou et al., 2012; Bent et al., 1995). Far upstream element-binding protein 2 (KHSRP) is also decreased in primiparous cells. This protein participates in RNA splicing and plays a role in the production of micro RNA that leads to gene regulation (Treiber et al., 2012). Though only three proteins in this category were identified, the indication from their regulation and the regulation of protein processing is a decrease in production of transcripts and proteins in the primiparous MEC relative to the virgin state.

3.4.2.5 Secreted proteins

Just as in the pregnant/virgin comparison, several typically secreted proteins were identified and decreased in abundance in the primiparous MEC relative to the virgin state of the cell. Albumin, AFP and TF were once again all identified as higher in the virgin. Since the same pattern is seen in the pregnant cells as the primiparous cells relative to virgin MEC, the question then shifts focus to why the virgin state might be displaying higher levels of these proteins, especially AFP which is supposed to be highest in pregnancy and is reported to have lasting effects on decreased risk of breast cancer (Vakharia and Mizejewski, 2000; Parikh et al., 2005). The decrease in TF might be explained in the same manner as it was for the pregnant state - iron is required as a cofactor during DNA replication and growth, which might not be occurring as often in primiparous MEC as in virgin MEC (Gomme and McCann, 2005). The abundance of albumin in the virgin state is still surprising and further investigations would help reveal whether this pattern of expression is consistent and served a purpose.

3.4.2.6 Other identified proteins

Voltage dependent anion channel 2 and IGHVDJ were both also down regulated in the primiparous MEC relative to the virgin MEC. Without more information about other related proteins, limited speculation can be made on their roles in the virgin cells, though the previous proposed explanation for decreased VDAC2 goes along with the seeming reduction in metabolism in the primiparous MEC relative to the virgin as well. A reduced metabolic load of the mitochondria may lack the need of more VDAC2 in the primiparous cells. One other protein of interest was not found as differing in abundance between virgin and pregnant states. Cell division cycle 42 (CDC42) is a protein that plays a role in cell cycle progression, cell polarity, migration and differentiation and has recently reported as required for MEC morphogenesis *in vitro* (Bray et al., 2011). This protein is also overexpressed in certain breast cancer types (Fritz et al., 2002). The down regulation of this protein from the virgin to the primiparous state of the MEC might be an interesting further investigation, as it has already been implicated in the possible role of hyper-proliferation and migration of cancer cells (Bray et al., 2013).

3.4.3 Comparison to Previous Studies

As mentioned previously, an investigation much like the one reported in this chapter has been performed in the same lab on isolated primary pregnant MEC compared with lactating MEC (Strand, 2012). Many identified proteins in the two comparisons were shared, though their regulation varied among them. The comparison between the pregnant and lactating states of the primary MECs showed largely up regulated metabolic processes. Many enzymes involved in fatty acid synthesis, the TCA cycle and the

respiratory chain were increased from the pregnant to the lactating state. This would be expected due to the high demand of the cell to produce and transport components into the alveolar lumen. An interesting comparison would be that of the virgin MEC and lactating MEC, as the pregnant MEC appeared to have decreased metabolic activity relative to the virgin cells. The down regulation of the cytoskeletal proteins, LASP1 and PDLIM1, was also seen in the previous pregnant/lactating comparison as it was in the virgin/pregnant comparison. This indicates that they are highest in the virgin state of the MEC, which may be an interesting topic of investigation due to the prior reports of high expression of these proteins relating to breast cancer (Johnsen et al., 2009; Lin et al., 2004). Patterns of protein synthesis and processing were similar to what was found in the HC11 *in vitro* investigation discussed in Chapter 2. Protein processing was down in lactating MEC even from pregnant MEC, which again indicates the high levels of protein synthesis and processing in the virgin MEC. A study examining the comparisons between many states of development in one data set would be very useful in finding changes and trends of protein category expression profiles among these MEC.

As mentioned in Chapter 1, D'Cruz and colleagues performed a microarray on parous vs. nulliparous rodent glands and found decreased expression of transcripts involved in proliferation and increased expression of genes involved in growth inhibition. Similar results were found by Ginger and colleagues when the expression profile of rats without exposure to gestational hormones was compared to rats that had previously been subjected to induced pregnancy through E and P levels (2001). In addition they found differences in metabolic genes and genes involved in cell-cell contact and the ECM. Balogh et al studied the breast tissue from nulliparous and parous women (2006). This

study also found evidence of increased immune surveillance and proliferation inhibition through apoptotic factors in parous breast tissue as well as changes in genes involved in cell trafficking, cell signaling, transcription and translation, and others. These investigations were performed on whole tissue from the breast, whereas the study recorded here was performed on isolated mammary epithelial cells. Though the possibility of contamination with other cell types of the mammary gland is not negligible, the data is likely more indicative of the molecular occurrences in that particular cell type than the others.

CHAPTER 4 – Development of a database for organization and analysis of mammary epithelial cell data: BROVINE

4.1 Introduction

In today's world of seemingly limitless ability for data generation from the scientific community, making sense of the vast sea of information is the largest obstacle. The ability of experts from different fields to communicate effectively and work together has become more important with this surge of data generation and the need for interdisciplinary communication is recognized and well documented (Wolkenhauer and Hofmeyr, 2013; NRC, 2010; Howe et al., 2008). Computer programs are needed for storing, organizing, and analyzing these large data sets and the scientists that produce this data are not likely capable of creating the advanced programs needed. The Peterson lab's investigation into the changing MEC through different stages of development has generated relatively small data sets that include lists of protein names that have changed in expression levels from one state to the next. The lists are growing and the desire to find relationships among the proteins on these different lists is a difficult achievement due the number of them present. A computer could accomplish this process much more quickly and easily than a human – if the right program existed.

The Peterson lab collaborated with Dr. Alex Dekhtyar from the computer science department in order to create a program that could help organize and analyze the data sets being produced. The aim of the database is to help unveil potential transcriptional regulators of mammary development by using lists of proteins, or their associated genes, that are differentially expressed between different stages of development in the MEC. Lists of potential transcription factors (TFs) for each differentially expressed gene are

generated with an online tool and this information needs to be managed as well. Ideally, the database would aid in upload and organization of data from many experiments performed in any species looking at any stage of mammary development. The database would then be responsible for organizing and analyzing these lists of genes and corresponding TFs to uncover relationships among species, genes, and the regulation patterns of these genes.

The following chapter may not be typical of a scientific paper as it is not filled with data sets and facts, rather a discussion regarding the personal experience of interdisciplinary learning and a tour of the end product. It discusses different steps taken and progress that has been made to begin development of software that can aid in the analysis of the MEC data sets being generated more quickly and in more depth than would otherwise be possible.

4.2 Methods

4.2.1 BIO 441

My participation in this class occurred after my participation in CSC 366 (described below), however, the information learned in this class would have been very useful prior to my involvement in CSC 366 so it will be explained first. The class was taught by Dr. Anya Goodman and included lecture and lab. It was split into two major learning objectives; students learned to annotate genes and participated in the Genomics Education Partnership by annotating a portion of the *Drosphilia melanogaster* genome, and they worked with computer science students to create computer programs that performed tasks related to genome assembly and analysis. The lab was typically held with computer science students from CS 448 taught by Dr. Alex Dekhtyar. The two classes were first taught basics of the others' discipline to provide background knowledge that would help with the collaboration that took place during lab sessions. Teams of BIO 441 students wrote program requirements outlining desired functionality for computer programs and CS 448 students wrote code to create said programs. As each portion was completed, the BIO 441 students would run tests and report back to the CS students in order to fix bugs found in the programs until their performance was satisfactory. The process of software development was introduced to BIO students in this way while CS students gained experience with a real client outside of the discipline.

4.2.2 CSC 366

To begin development of the database desired for the Peterson lab's growing data set, we became the clients for Dr. Alex Dekhtyar's class of computer science students (CSC 366) whose job was to produce our desired product. The class began with a brief lesson on genetics and transcriptional regulation to give the CS students a background for the database needed for our project. In order for our desired program plan to be conveyed properly to the CS students, terms needed to be defined and exact functions needed to be specified so the developers knew what code they needed to write; we constructed a document containing use cases, which outlined the steps describing our desired interactions with the program. The document outlined specifics regarding data insertion and manipulation functionality, browsing and filtering functionality, and analytical functionality (see Appendix A). From that point the class broke into teams and each team was in charge of developing prototypes of the product we requested. The developers used our document to formulate requirements and begin to design the software. As portions of the database were implemented they were tested with initial data we provided. Adjustments were made based off of bugs found during testing and clarifications of desired functionalities. At the end of the class, the complete database products were presented to us and one was chosen for further development. The database chosen was web-based and was named "Brovine" by the developing team. It was developed by a team of men and it was developed using data from a bovine cell line, which led to the developers naming the program "Brovine."

4.2.3 Individual database development

The desired functionality of the database went beyond the scope of the CSC 366 class. One team member, Therin Irwin, that helped create Brovine continued development of the database through individual consultation with the Peterson Lab as his senior project. Further improvements were implemented in a similar manner. Documents similar in form to the use case document presented in Appendix A were written containing desired functionalities. Requirements were then formulated and the new code was written. The new portions of the program were then tested by the Peterson lab and a report was sent back with any bugs detected or changes desired.

4.2.4 Data Preparation

The data used to develop the database was generated by a 2DGE experiment performed by former lab mate, Laura Strand. A bovine cell line, MAC-T was cultured and protein was extracted and analyzed in the undifferentiated and differentiated states. Differentially expressed proteins between the two states were identified and a list was generated. Promoter regions for the genes of this list of differentially expressed proteins were obtained from their associated gene sequence found on the NCBI genome database, starting 1950 nucleotides before the transcription start site and ending 50 nucleotides after the transcription start site. Lists of transcription factors associated with each of these genes were generated using Transcription Element Search System (TESS). TESS was an available online tool that identified potential transcription factors for a gene based on presence of the TF's regulatory elements in a provided sequence. The resulting product was an excel file containing worksheets listing the search parameters used and the potential TFs associated with the provided promoter sequence. These excel files were then converted to .csv files to allow loading into the developing databases in CSC 366.

4.3 Results

At the end of the CSC 366 class, all use case specifications (Appendix A) excluding the extra credit items were completed by many of the developed databases including the chosen database, Brovine. These included pages labeled Upload, Experiment Hierarchy, Gene Summary, TF Summary, Gene Search, and TF Popularity.

After continued development with Therin Irwin, existing pages were updated further and a page labeled TF Subtract was added. The following were the final pages containing the most important functionalities and their descriptions:

1. Upload – The upload page of the Brovine database allows individual gene or bulk file upload into the web-based database. The user is able to browse the files on their computer to upload into the database. Files are uploaded in .csv format, which are converted from excel files. A snapshot of this page is shown in Figure 7.

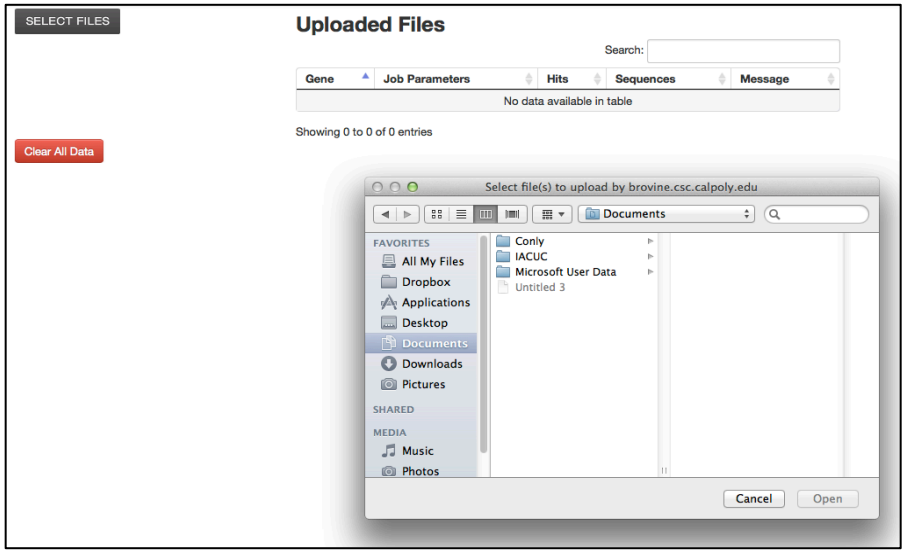

Figure 7. *Upload page of the Brovine database*.

2. Experimental Hierarchy – The experimental hierarchy page allows the used to filter down through the uploaded data to find information on specific known entities. The flow of the information begins with a choice of species, followed by comparison, followed by experiment, followed by gene, followed by transcription factor, followed by regulatory sequence. When a regulatory sequence is selected, more information about it is displayed including the start position, its length, the sense, and the exact nucleotide sequence. This page also allows editing of the information that was uploaded as well as the ability to hide information. A snapshot of this page is shown in Figure 8.

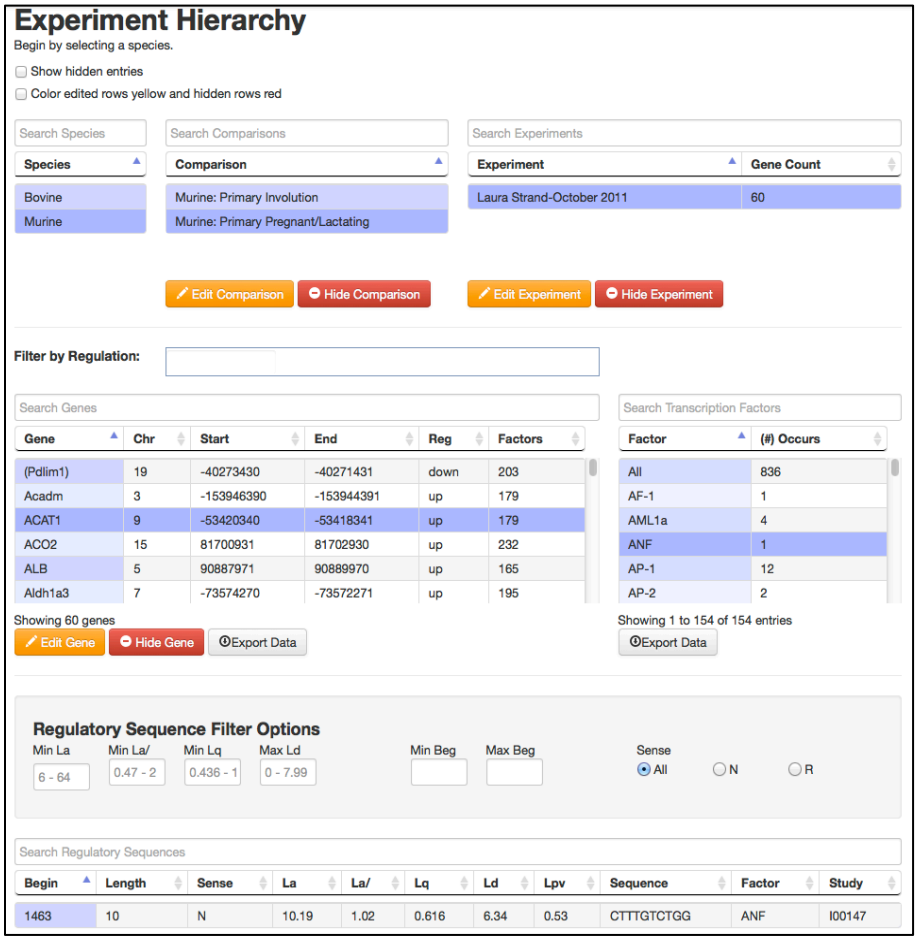

Figure 8. *Experiment Hierarchy page of the Brovine database*.

3. Gene Summary – This page provides a list of all genes populating the database in a table form. Information provided in the table includes the gene name and abbreviation, the number of comparisons in which the gene is found, and the number of experiments in which the gene is found. The table is sortable by any column. When a gene is selected, another table appears listing the experiments in which the gene is found and the following information found in that experiment: the corresponding comparison and species, the gene's regulation in that experiment, and the position of the gene in that species' genome. The second table listing the experiments can be filtered by regulation. A snapshot of this page is shown in Figure 9.

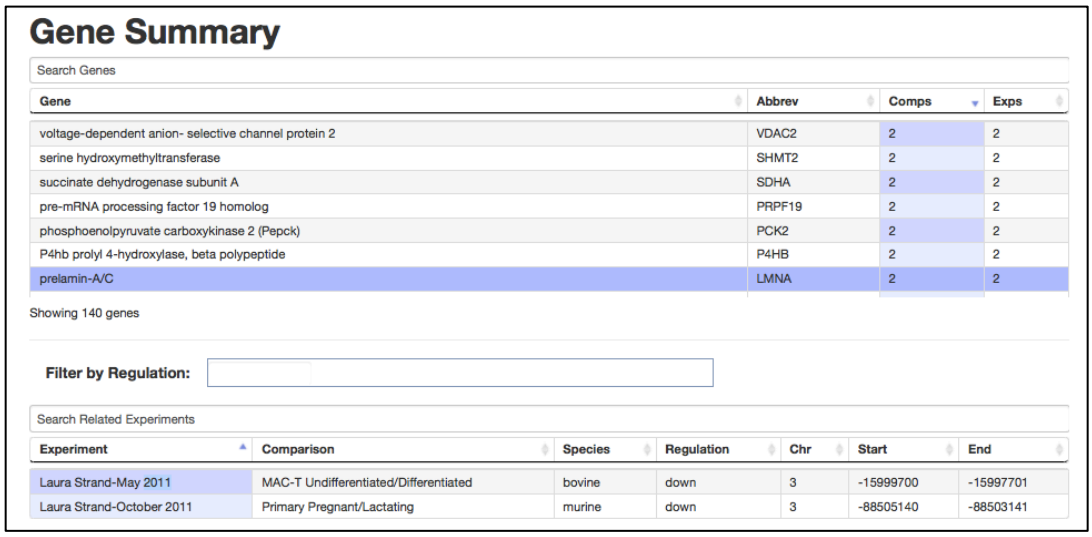

Figure 9. *Gene Summary page of the Brovine database*.

4. TF Summary – This page provides a list of all transcription factors populating the database in a table form. Information provided in the table includes the TF name, the number of genes in which the TF is found, and the total number of occurrences of the TF in the database. The table is sortable by any column. When a TF is selected, another table appears listing the occurrences and including the following information about each one: the length and sense and beginning position of the sequence, the species, comparison and experiment the occurrence exists, the gene in which it was found and that gene's regulation in the corresponding experiment. The second table listing the occurrences can be filtered by regulation. A snapshot of this page is shown in Figure 10.

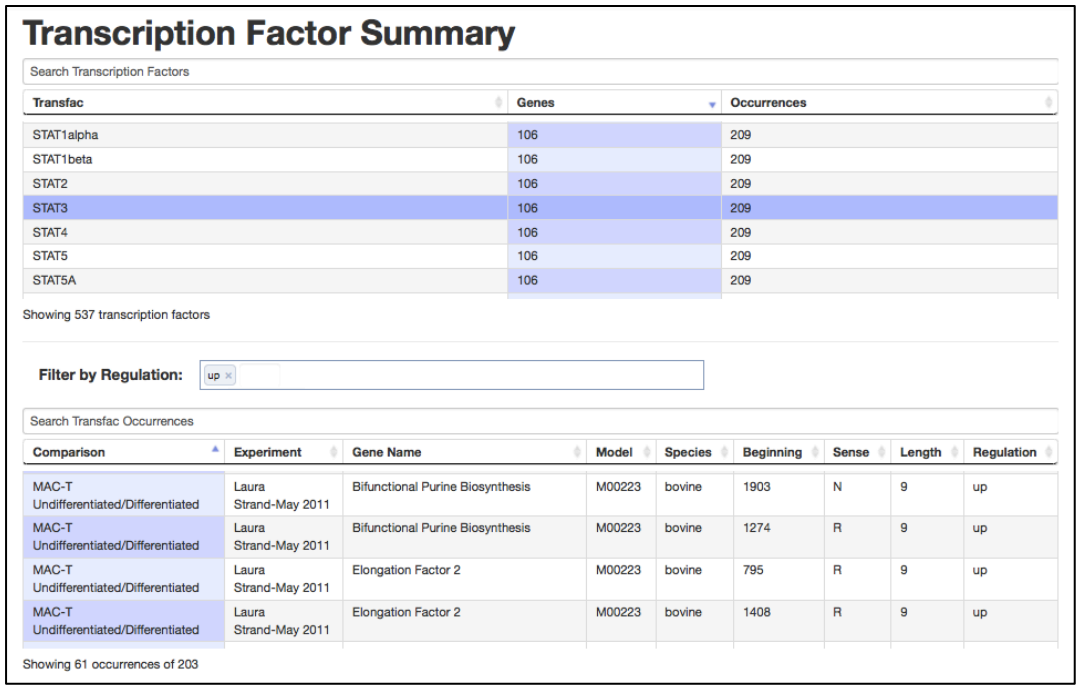

Figure 10. *TF Summary page of the Brovine database*.

5. Gene Search – This page allows the user to select any number species, comparisons and experiments to provide a list of TF present in the selected categories. The user can then select any number of specific TFs from this list and specify filtering options. The database will then use these selections to provide a list of genes containing all of the selected TFs that meet the filtering requirements. A snapshot of this page is shown in Figure 11.

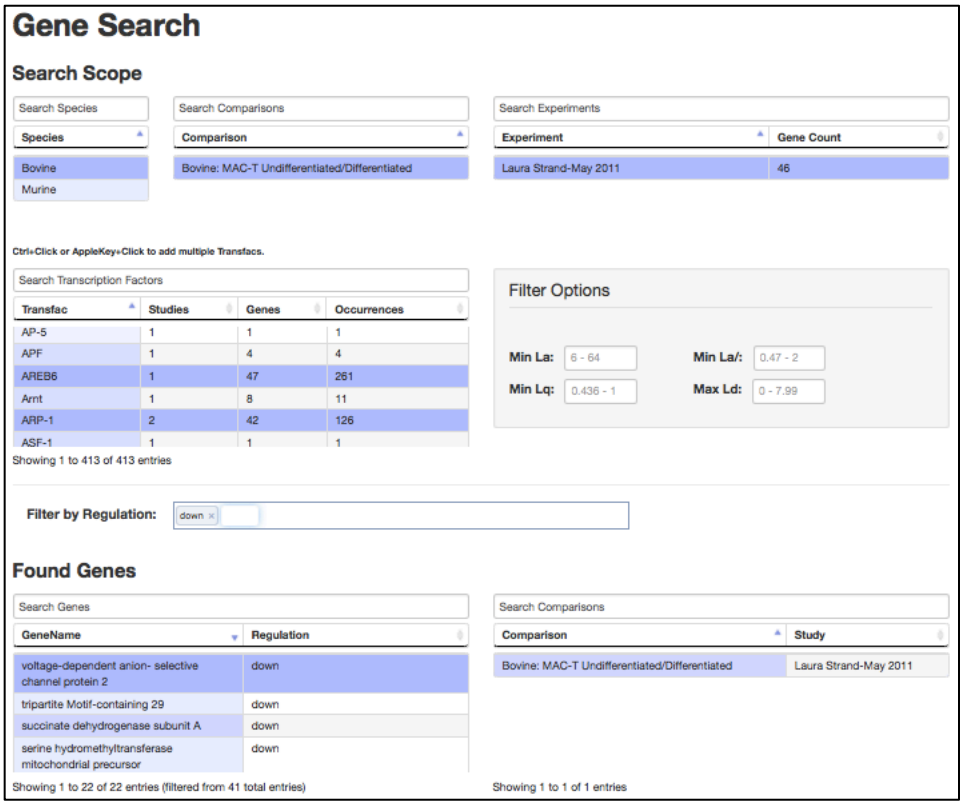

Figure 11. *Gene Search page of the Brovine database*.

6. TF Subtract – This page allows the user to select two separate groups of genes and subtract the TFs present in one group of genes from the list of TFs present in the other group of genes. This allows the user to find TFs that might have potential significance in an experimental group of differentially expressed genes by showing only the TFs that are present in the experimental comparison, but not in a control group of genes. The final list provided includes the TF names, the number of genes in which they appear, and their number of occurrences. A snapshot of this page is shown in Figure 12.

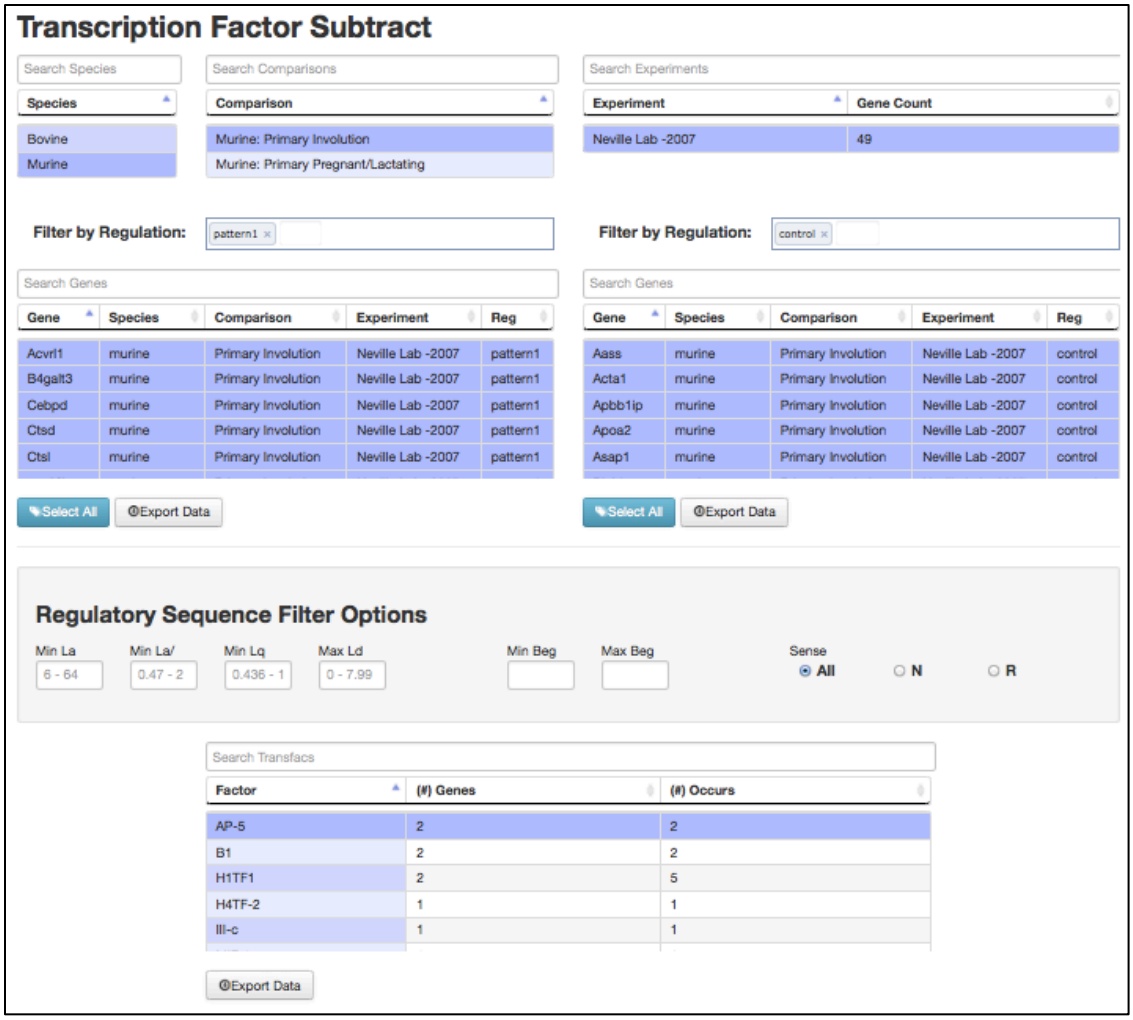

Figure 12. *TF Subtract page of the Brovine database*.

4.4 Discussion

Development of the database is not complete, but much has been accomplished. No publishable experimental results have been obtained from the database to this point. Further functionality is desired from this web-based program related to time course studies and regulation of genes sets throughout the duration of them. The TF Subtract page was a large advancement in this direction, but eventual development will aim at creating regulation profiles for different gene transcripts found at different time points of mammary development, grouping these genes based on their profiles, then using the functionality of the TF subtract page in order to find TFs that could potentially be driving the regulatory profiles of these groups of genes. When a few key TFs are found by the database that have potential of driving regulatory mechanisms of mammary development, experiments can be performed to confirm or refute these factors' role in development.

Other groups have designed software tools similar to the one we are attempting to generate. A description of other, similar tools follows. *TFM-Explorer* has the goal of identifying over-representation of transcription factor binding sites within a set of user-defined genes (Tonon et al., 2010). This seems to be a useful tool for finding over-represented transcription factors that works on the basis of probabilities of sequence presence, but does not appear to have a subtraction of a control set. Without the subtraction it would be hard to determine which transcription factors are important in solely the cell model in which the investigator is interested or in many models. *CORE_TF* performs a search for over-represented transcription factor binding sites based on user-provided ENSEMBL gene IDs and user-defined promoter regions (Hestand et al., 2008). It presents a rather overwhelming table listing all identified binding sites in both

the control set of genes and the users' genes of interest. This tool has a similar first goal of our database by finding binding sites taking into account a control set and has the ability of finding homologous genes. *CARRIE* is a tool that accomplishes goals very similar to ours, allowing the user to input microarray data and then proceeding to determine gene expression changes and identify likely transcription factor binding sites regulating the co-regulation of these genes. It also goes to the next step to construct regulatory networks, as is one of our final goals (Haverty et al., 2004). This database seems to be a great model for what we would like to accomplish, but with updated data input capabilities, as RNAseq is the new way to analyze gene expression, and additional ability to identify promoter patterns as mentioned above. We can use the lessons provided by these other tools to improve on their functionality in ours.

The interdisciplinary collaboration experienced throughout the development of this database was extremely productive for all parties involved. The computer scientists and animal scientists were able to teach one another about their respective fields and the difficulties that can arise in the communication between the two disciplines was realized and eventually overcome. The Peterson lab was exposed to the process of software development and learned how to effectively communicate software requirements to the developers, including the level of specificity required. The computer scientists learned the importance of understanding the information they are working with and the basics of the field for which they are developing software. This real world experience should benefit all parties by giving them insight into the growing reality of interdisciplinary studies and the necessary collaboration between the technical and science fields.

LIST OF REFERENCES

Abdul-Awal, et al., 1996. Ultrastructural and morphometrical studies on the endothelial cells of arteries supplying the abdomino-inguinal mammary gland of rats during the reproductive cycle. J Vet Med Sci 58(1):29-34.

Alberts, et al., (eds.), 2002. Cell Junctions, Cell Adhesion, and the Extracellular Matrix. In *Molecular Biology of the Cell. 4th edition*. New York: Garland Science.

Anderson, et al., 2007. Secretory activation in the mammary gland: it's not just about milk protein synthesis! Breast Cancer Res 9(1):204.

Andres and Djonov, 2010. The mammary gland vasculature revisited. J Mammary Gland Biol Neoplasia 15:319-328.

Atabai, et al., 2007. Roles of the innate immune system in the mammary gland remodeling during involution. J Mammary Gland Biol Neoplasia 12(1):37-45.

Auperlee, et al., 2013. Amphiregulin mediates progesterone-induced mammary ductal development during puberty. Breast Cancer Res 15(3):R44.

Ball, et al., 1988. Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. EMBO J 7(7):2089-2095.

Balogh, et al., 2006. Genomic signature induced by pregnancy in the human breast. Int J Oncol 28:399-410.

Bartek et al., 1990. Keratin 19 expression in the adult and developing human mammary gland. Histochem J 22:537-544.

Bauman and Vernon, 1993. Effects of exogenous bovine somatotropin on lactation. Annu Rev Nutr 13:437–461.

Beleut, et al., 2010. Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland. PNAS 107(7):2989-2994.

Bent, et al., 1995. Heterogenous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. J Biol Chem 270:28780-28789.

Blachford, et al., 1995. Vectorial secretion by constitutive and regulated secretory pathways in mammary epithelial cells. Epith Cell Biol 4:8-16.

Blaser, et al., 2003. Isolation of new splice isoforms, characterization and expression analysis of the human septin *SEPT8*. Gene 312:313-320.

Boumahrou, et al., 2009. The major protein fraction of mouse milk revisited using proven proteomic tools. J Physiol Pharmacol 60(Suppl 3):113-118.

Bourges, et al., 2008. New insights into the dual recruitment of IgA+ B cells in the developing mammary gland. Mol Immunol 45:3354-3362.

Bray, et al., 2011. The Rho GTPase Cdc42 is required for mammary epithelial cell morphogenesis in vitro. Small GTPases 2:247–258.

Bray et al., 2013. Cdc42 overexpression induces hyperbranching in the developing mammary gland by enhancing cell migration. Breast Cancer Res 15:R91.

Brisken, et al., 1998. A paracrine role for the epithelial progesterone receptor in mammary gland development. Proc. Natl. Acad. Sci. 95: 5076–5081.

Brisken, et al., 1999. Prolactin controls mammary gland development
via direct and indirect mechanisms. Dev Biol 210:96-106.

Brisken, et al., 2000. Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. Genes Dev 14(6):650-654.

Brisken, et al., 2002. IGF-2 Is a mediator of prolactin-induced morphogenesis in the breast. Dev Cell 3(6):877-887.

Brisken, 2002. Hormonal control of alveolar development and its implications for breast carcinogenesis. J Mammary Gland Biol Neoplasia 7(1):39-48.

Britt, et al., 2007. Pregnancy and the risk of breast cancer. Endocr Relat Cancer 14:907-933.

Buratte and Buralle, 2008. Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. Front Biosci 13:867-878.

Butcher, et al., 2009. A tense situation: Forcing tumour progression. Nat Rev Cancer 9(2):108-122.

Butte, et al., 1999. Adjustments in energy expenditure and substrate utilization during late pregnancy and lactation. Am J Clin Nutr 69:299–307.

Byrne, et al., 2000. Identification of a hormone-responsive promoter immediately upstream of exon 1c in the human vitamin D receptor gene. Endocrinology 141:2829–36.

Capuco and Ellis, 2005. Bovine mammary progenitor cells: Current concepts and future directions. J Mammary Gland Biol Neoplasia. 10(1):5-15.

Capuco, et al., 1999. Prolactin and growth hormone stimulation of lactation in mice requires thyroid hormones. Proc Soc Exp Biol Med 221(4):345-351.

Cardiff and Allison, 2012. Mammary Gland. In Treuting and Dintzis (eds.), *Comparative Anatomy and Histology: A Mouse and Human Atlas*. Academic Press.

Chen, et al. 2005. The HSP90 family of genes in the human genome: Insights into their divergence and evolution. Genomics 86:627-637.

Chen, et al., 2010. Akt is required for Stat5 activation and mammary differentiation. Breast Cancer Res 12:R72.

Chen, et al., 2012. Autocrine prolactin induced by the Pten-Akt pathway is required for lactation initiation and provides a direct link between the Akt and Stat5 pathways. Genes Dev 26:2154-2168.

Choi, et al., 2009. Elf5 conditional knockout mice reveal its role as a master regulator in mammary alveolar development: Failure of Stat5 activation and functional differentiation in the absence of Elf5. Dev Biol 329:227-241.

Cid, et al., 2002. Estrogens and the vascular endothelium. Ann N Y Acad Sci 966:143–157.

Clarke, et al., 1997. Dissociation between steroid receptor expression and cell proliferation in the human breast. Cancer Res 57:4987-4991.

Clegg, 1981. Lipoprotein lipase. Localization on plasma membrane fragments from lactating rat mammary tissue. Biochim Biophys Acta 664(2):397–408.

Collier, et al., 1977. Lactogenesis in explant cultures of mammary tissue from pregnant cows. Endocrinology 100(4):1192-1200.

Creamer, et al., 2010. Stat5 promotes survival of mammary epithelial cells through transcriptional activation of a distinct promoter in Akt1. Mol Cell Biol 30(12):2957-2970.

D'Cruz, et al., 2002. Persistent parity-induced changes in growth factors, TGF-beta 3, and differentiation in the rodent mammary gland. Mol Endocrinol 16(9):2034-2051.

Daniel and Silberstein, 1987. Postnatal development of the rodent mammary gland. In Neville and Daniel, (eds.), *The Mammary Gland: Development, Regulation,*

and Function. New York: Plenum Press.

Daugaard, et al., 2007. The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. FEBS Lett 581:3702-3710.

Desrivieres, et al., 2003. Comparative proteomic analysis of proliferating and functionally differentiated mammary epithelial cells. Mol Cell Proteomics 2:1039-1054.

Djonov, et al., 2001. Vascular remodelling during the normal and malignant life cycle of the mammary gland. Microsc Res Tech 52:182-189.

Dreyfuss, et al., 2002. Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol 3(3):195-205.

Edery, et al., 1985. Regulation of estrogen and progesterone receptor levels in mouse mammary epithelial cells grown in serum-free collagen gel cultures. Endocrinology 116:105–112.

Elias et al., 1973. Changes in fat cell morphology during lactation in the mouse. Anat Rec 177:533-548.

Emerman and Vogl, 1986. Cell size and shape changes in the myoepithelium of the mammary gland during differentiation. Anat Rec 216:405-415.

Fata, et al., 2000. The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. Cell 103:41-50.

Faulkner and Peaker, 1987. Regulation of mammary glucose metabolism in lactation. In Neville and Daniel (eds.), *The Mammary Gland: Development, Regulation and Function*. New York: Plenum.

Fendrick, et al., 1998. Mammary gland growth and development from the postnatal

period to postmenopause: Ovarian steroid receptor ontogeny and regulation in the mouse. J Mammary Gland Biol Neoplasia 3(1):7-22.

Feng, et al., 1995. Glucocorticoid and progesterone inhibit involution and programmed cell death in the mouse mammary gland. J Cell Biol 131:1095–1103.

Fernandez-Valdivia, et al., 2009. The RANKL signaling axis is sufficient to elicit ductal side-branching and alveologenesis in the mammary gland of the virgin mouse. Dev Biol 328:127-139.

Feuerrman et al., 2009. Mammary fat can adjust prolactin effect on mammary epithelial cells via leptin and estrogen. Int J Endocrinol 2009:427260.

Flint and Vernon, 1998. Effects of food restriction on the responses of the mammary gland and adipose tissue to prolactin and growth hormone in the lactating rat. J Endocrinol 156:299–305.

Freund, et al., 2012. Lamin B1 loss is a senescence-associated biomarker. Mol Biol Cell 23(11):2066-2075.

Fritz, et al., 2002. Rho GTPases in human breast tumours: expression and mutation analyses and correlation with clinical parameters. Br J Cancer 87:635–644.

Ginger, et al., 2001. Persistent changes in gene expression induced by estrogen and progesterone in the rat mammary gland. Mol Endocrinol 15(11):1993-2009.

Goldman, et al., 2002. Nuclear lamins: building blocks of nuclear architecture. Genes Dev 16:533-547.

Gomme and McCann, 2005. Transferrin: structure, function and potential therapeutic actions. Drug Discov Today 10(4):267-73.

Gooding, et al., 2010. Three-dimensional ultrasound imaging of mammary ducts in

lactating women: a feasibility study. J Ultrasound Med 29(1):95-103.

Gors, et al., 2009 Technical note: Milk composition in mice - methodological aspects and effects of mouse strain and lactation day. J Dairy Sci 92(2):632-637.

Gouon-Evans, et al., 2000. Postnatal mammary gland development requires macrophages and eosinophils. Development 127:2269-2282.

Gouon-Evans, et al., 2002. Requirement of macrophages and eosinophils and their cytokines/chemokines for mammary gland development. Breast Cancer Res 4:155-164.

Hadsell and Bonnette, 2000. IGF and insulin action in the mammary gland: Lessons from transgenic and knockout models. J Mammary Gland Biol Neoplasia 5:19–30.

Hassiotou and Geddes, 2013. Anatomy of the human mammary gland: Current status of knowledge. Clin Anat 26(1):29-48.

He and Smith, 2009. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. Cell Mol Life Sci 66:1239-1256.

Hennighausen and Sippel, 1982. Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. Nucleic Acids Res 10(8):2677-2684.

Mouse whey acidic protein is a novel member of the family of ‘four-disulfide core’ proteins. Nucl Acids Res 10(8):2677-2684.

Hollman, 1974. Cytology and fine structure of the mammary gland. In Larson and Smith, (eds.), *Lactation*. New York: Academic Press.

Hovey and Aimo, 2010. Diverse and active roles for adipocytes during mammary gland growth and function. J Mammary Gland Biol Neoplasia 15:279-290.

Hovey, et al., 1999. Regulation of mammary gland growth and morphogenesis by the mammary fat pad: A species comparison. J Mammary Gland Biol Neoplasia 4(1):53-68.

Hovey, et al., 2001. Transcriptional regulation of vascular endothelial growth factor expression in epithelial and stromal cells during mouse mammary gland development. Mol Endocrinol 15(5):819-831.

Hovey, et al., 2003. Local insulin-like growth factor-II mediates prolactin-induced mammary gland development. Mol Endocrinol 17(3):460-471.

Howard and Gusterson, 2000. Human breast development. J Mammary Gland Biol Neoplasia 5(2):119-137.

Howe, et al., 2008. Big data: The future of biocuration. Nature 455(7209):47-50.

Howlin, et al., 2006. Pubertal mammary gland development: Insights from mouse models. J Mammary Gland Biol Neoplasia 11:283-297.

Humphreys, et al., 1996. Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis. Development 122:4013-4022.

Hunziker and Kraehenbul, 1998. Epithelial transcytosis of immunoglobulins. J Mammary Gland Biol Neoplasia 3(3):287-302.

Ingle, et al., 1972. Lipogenesis in the ruminant: in vitro study of tissue sites, carbon source and reducing equivalent generation for fatty acid synthesis. J Nutr 102:609-616.

Ito, et al., 2009. Sept8 controls the binding of vesicle-associated membrane protein 2 to synaptophysin. J Neurochem 108:867-880.

Jenness, 1974. Biosynthesis and composition of milk. J Invest Dermatol 63:109-118.

Jenness, 1988. Composition of milk. In Wong, et al., (eds.), *Fundamentals of Dairy Chemistry*. Springer US.

Jensen, et al., 1994. Regulation of lipoprotein lipase activity and mRNA in the mammary gland of the lactating mouse. Biochem J 298:321-327.

Jensen, (ed.) 1995. *Handbook of milk composition*. San Diego: Academic Press.

Johnsen, et al., 2009. Regulation of estrogen-dependent transcription by the LIM cofactors CLIM and RLIM in breast cancer. Cancer Res 69:128-136.

Johnson, et al., 2001. Limits to sustained energy intake. I. Lactation in the laboratory mouse Mus musculus. J Exp Biol 204(Pt 11):1925-35.

Kabmeyer, et al., 2009. New insights in vascular development: Vasculogenesis and endothelial progenitor cells. Anat Histol Embryol 38:1-11.

Kalsheker, 1988. Alpha1-antitrypsin: Structure, function and molecular biology of the gene. Biosci Rep 9(2):129-138.

Kamada and Miwa, 1992. A protein binding to CArG box motifs and to single-stranded DNA functions as a transcriptional repressor. Gene 119(2):229-236.

Karg and Schams, 1974. Prolactin release in cattle. J Reprod Fert 39:463-472.

Kelsey, et al., 1993. Reproductive factors and breast cancer. Epidemiol Rev 15(1):36-47.

Kingsley-Kallesen, et al., 2002. The mineralocorticoid receptor may compensate for the loss of the glucocorticoid receptor at specific stages of mammary gland development. Mol Endocrinol 16:2008–2018.

Kinsella and Whitehead, 1989. Proteins in whey: Chemical, physical, and functional properties. Adv Food Nutr Res 33.3: 437-438.

Kleinberg, 1997. Early mammary development: Growth hormone and IGF-1. J Mammary Gland Biol Neoplasia 2(1):49-57.

Knight and Beever, 1994. Nutrient metabolism and utilization in the mammary gland. Livestock Prod Sci 39:129-137.

Koll, et al., 1992. Antifolding activity of hsp60 couples protein import into the mitochondrial matrix with export to the intermembrane space. Cell 68:1163-1175.

Koya, et al., 2000. Gelsolin inhibits apoptosis by blocking mitochondrial membrane potential loss and cytochrome c release. J Biol Chem 275:15343-15349.

Koyama, et al., 2013. The nipple: A simple intersection of mammary gland and integument, but focal point of organ function. J Mammary Gland Biol Neoplasia 18:121-131.

Kuhn, et al., 1980. Lactose synthesis: The possibilities of regulation. J Dairy Sci 63:328-336.

Lambe, et al., 1994. Transient increase in the risk of breast cancer after giving birth. N Engl J Med 331:5-9.

Lambe, et al., 1996. Parity, age at first and last birth, and risk of breast cancer: A population-based study in Sweden. Breast Cancer Res Treat 38(3):305-11.

Lane-Clayton, 1926. A further report on cancer of the breast, with special reference to its associated antecedent conditions. In *Rep Min Health No. 32*. London: HM Stationery Office.

Lascelles and Lee, 1978. Involution of the mammary gland. In Larson and Smith (eds.), *Lactation - A Comprehensive Treatise, Vol. 4*. New York: Academic Press.

Lattanzi, 2011. Prelamin A-mediated nuclear envelope dynamics in normal and laminopathic cells. Biochem Soc Trans 39:1698-1704.

Lee, et al., 2013. Progesterone drives mammary secretory differentiation via RankL-mediated induction of Elf5 in luminal progenitor cells. Development 140:1397-1401.

Lilla and Werb, 2010. Mast cells contribute to the stromal microenvironment in mammary gland branching morphogenesis. Dev Biol 337:124-133.

Lin, et al., 2004. Regulation of cell migration and survival by focal adhesion targeting of Lasp-1. J Cell Biol 165:421-432.

Liu, et al., 1997. StatSa is mandatory for adult mammary gland development and lactogenesis. Genes Dev 11:179-186.

Lonnerdal, 1985. Biochemistry and physiological function of human milk proteins. Am J Clin Nutr 42:1299-1317.

Luhr, et al., 2012. Mammary fibroblasts regulate morphogenesis of normal and tumorigenic breast epithelial cells by mechanical and paracrine signals. Cancer Lett 325:175-188.

Lund, et al., 1996. Two distinct phases of apoptosis in mammary gland involution: proteinase-independent and -dependent pathways. Development 122:181-193.

Ma, et al., 2002. Heterogeneous nuclear ribonucleoprotein A3, a novel RNA trafficking response element-binding protein. J Biol Chem 277:18010-18020.

Ma, et al., 2006. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. Breast Cancer Res 8:R43.

MacMahon, et al., 1970. Age at first birth and breast cancer risk. Bull World Health Organ 43(2):209-21.

Mallepell, et al., 2006. Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. PNAS 103(7):2196-2201.

Maller, et al., 2010. Extracellular matrix composition reveals complex and dynamic stromal-epithelial interactions in the mammary gland. J Mammary Gland Biol Neoplasia 15:301-318.

May, 2013. Engaging students in a bioinformatics activity to introduce gene structure and function. J Micobiol Biol Educ 14(1):107-109.

McCave, et al., 2010. The normal nicroenvironment directs mammary gland development. J Mammary Gland Biol Neoplasia 15:291-299.

Mepham, (ed.), 1983. *Biochemistry of lactation*. Amsterdam; New York: Elsevier.

Mezl and Knox, 1977. Metabolism of arginine in lactating rat mammary gland. Biochem J 164:105-113.

Mikaelian, et al., 2006. Expression of terminal differentiation proteins defines stages of mouse mammary gland development. Vet Pathol 43:36-49.

Moriera, et al., 2010. Tissue proteomics of the human mammary gland: Towards an abridged definition of the molecular phenotypes underlying epithelial normalcy. Mol Oncol 4:539-561.

Morrison and Cutler, 2009. Mouse mammary epithelial cells form mammospheres during lactogenic differentiation. J Vis Exp (32).

Naccarato, et al., 2003. Definition of the microvascular pattern of the normal human adult

mammary gland. J Anat 203:599-603.

Nagato, et al., 1980. A scanning electron microscope study of myoepithelial cells in exocrine glands. Cell Tissue Res 209:1-10.

Nemeth, et al., 2000. Golgi targeting of the GLUT1 glucose transporter in the lactating mouse mammary gland. Pediatr Res 47:444-450.

Neville and Daniel, (eds.), 1987. *The mammary gland: Development, regulation and function*. New York: Plenum Press.

Neville and Picciano, 1997. Regulation of milk lipid secretion and composition. Annu Rev Nutr 17:159-184.

Neville, et al., 1998. The mammary fat pad. J Mammary Gland Biol Neoplasia 3(2):109-116.

Neville, et al., 2001. Lactogenesis. The transition from pregnancy to lactation. Pediatr Clin North Am 48(1):35-52.

Neville, et al., 2002. Hormonal regulation of mammary differentiation and milk secretion. J Mammary Gland Biol Neoplasia 7(1):49-66.

Nguyen and Neville, 1998. Tight junction regulation in the mammary gland. J Mammary Gland Biol Neoplasia 3(3):233-246.

National Research Council (NRC), 2010. Bio2010: Transforming Undergraduate Education for Future Research Biologists. Washington, DC: The National Academies Press.

Oakes, et al., 2006. The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. Breast Cancer Res 8(2):207.

Oakes, et al., 2008. The Ets transcription factor Elf5 specifies mammary alveolar cell fate. Genes Dev 22(5):581-586.

Oka and Perry, 1974. Arginase affects lactogenesis through its influence on the biosynthesis of spermidine. Nature 250:660-661.

Ormandy, et al., 1997. Mammary gland development in prolactin receptor knockout mice. J Mamm Gland Biol Neoplasia 2:355-364.

Palmiter, 1969. Hormonal induction and regulation of lactose synthetase in mouse mammary gland. Biochem J 113:409-417.

Papadopoulou, et al., 2012. Expression profile and interactions of hnRNP A3 within hnRNP/mRNP complexes in mammals. Arch Biochem Biophys 523:151-160.

Parikh, et al., 2005. Prevention of N-methyl-N-nitrosourea-induced breast cancer by alpha-fetoprotein (AFP)-derived peptide, a peptide derived from the active site of AFP. Clin Cancer Res 11:8512-8520.

Patel and Korotchkina, 2006. Regulation of the pyruvate dehydrogenase complex. Biochem Soc Trans 34:217-222.

Pellegrino, et al., 1988. Differential expression of keratins 13 and 16 in normal epithelium, benign lesions, and ductal carcinomas of the human breast determined by the monoclonal antibody Ks8.12. Cancer Res 48:5831-5836.

Plant and Barker-Gibbs, 2004. Neurobiological mechanisms of puberty in higher primates. Hum Reprod Update 10(1):67-77.

Pollard and Henninghausen, 1994. Colony stimulating factor 1 is required for mammary gland development during pregnancy. Proc Natl Acad Sci 91:9312-9316.

Porter, et al., 2000. K15 expression implies lateral differentiation within stratified

epithelial basal cells. Lab Invest 80(11):1701-1710.

Quarrie, et al., 1995. Apoptosis in lactating and involuting mouse mammary tissue demonstrated by nick-end DNA labelling. Cell Tisssue Res 281:413-419.

Rabbany, et al., 2003. Molecular pathways regulating mobilization of marrow-derived stem cells for tissue revascularization. Trends Mol Med 9(3):109-117.

Ramamoorthy, et al., 2010. Isoform-specific degradation of PR-B by E6-AP is critical for normal mammary gland development. Mol Endocrinol 24:2099–2113.

Ramsay, et al., 2005. Anatomy of the lactating human breast redefined with ultrasound imaging. J Anat 206:525-534.

Reece, 2009. Lactation. In *Funcitonal Anatomy and Physiology of Domestic Animals*. John Wiley & Sons.

Richert, et al., 2000. An atlas of mouse mammary gland development. J Mammary Gland Biol Neoplasia 5(2):227-241.

Rosen, et al., 1999. Regulation of milk protein gene expression. Annu Rev Nutr 19:407–436.

Rostovtseva and Bezrukov, 2008. VDAC regulation: Role of cytosolic proteins and mitochondrial lipids. J Bioenerg Biomembr 40(3):163-170.

Ruan, et al., 1995. Estradiol enhances the stimulatory effect of insulin-like growth factor-1 (IGF-1) on mammary development and growth hormone induced IGF-1 messenger ribonucleic acid. Endocrinology 136:1296-1302.

Rudolph, et al., 2007. Metabolic regulation in the lactating mammary gland: A lipid synthesizing machine. Physiol Genomics 28:323-336.

Rudolph, et al., 2011. Prolactin-mediated regulation of lipid biosynthesis genes in vivo in the lactating mammary epithelial cell. Am J Physiol Endocrinol Metab 300:E1059-E1068.

Russo, et al., 1999. Pattern of distribution of cells positive for estrogen receptor alpha and progesterone receptor in relation to proliferating cells in the mammary gland. Breast Cancer Res Treat 53(3):217-227.

Sadaie et al., 2013. Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. Genes Dev 27:1800-1808.

Sakai, et al., 1979. Glucocorticoid regulation of prolactin receptors on mammary cells in culture. Endocrinology 104:1447–1449.

Salway, 1994. *Metabolism at a glance*. Blackwell Publishing, Ltd.

Sans, et al., 2006. MondoA-Mlx heterodimers are candidate sensors of cellular energy status: Mitochondrial localization and direct regulation of glycolysis. Mol Cell Biol 26(13):4863-4871.

Schedin, et al., 2000. Estrous cycle regulation of mammary epithelial cell proliferation, differentiation, and death in the Sprague-Dawley rat: A model for investigating the role of estrous cycling in mammary carcinogenesis. J Mammary Gland Biol Neoplasia 5(2):211-225.

Shi, et al., 2004. Hormonal defect in maspin heterozygous mice reveals a role of progesterone in pubertal ductal development. Mol Endocrinol 18(9):2196-2207.

Shiryaev, et al., 2012. Septin 8 is an interaction partner and in vitro substrate of MK5. World J Biol Chem 3(5):98-109.

SILS, 2013. “Bioinformatics – Academic Degree Programs.” *Informatics Degree Program Summary. UNC – School of Information and Library Science*. Web. 11 October 2013. <http://ils.unc.edu/informatics_programs/index.php>.

Simian, et al., 2001. The interplay of matrix metalloproteinases, morphogens and growth factors is necessary for branching of mammary epithelial cells. Development 128(16):3117-3131.

Srivastava, et al., 2003. Receptor activator of NF-kB ligand induction via Jak2 and Stat5a in mammary epithelial cells. J Biol Chem 278:46171-46178.

Stark, et al., 1997. Visualization of elongation factor Tu on the *Escherichia coli* ribosome. Nature 389(6649):403-406.

Stein, et al., 2007. Mammary gland involution as a multi-step process. J Mammary Gland Biol Neoplasia 12:25-35.

Stelwagen, et al., 2009. Immune components of bovine colostrum and milk. J Anim Sci 87:3-9.

Sternlicht, et al., 2005. Mammary ductal morphogenesis requires paracrine activation of stromal EGFR via ADAM17-dependent shedding of epithelial amphiregulin. Development 132(17):3923-3933.

Sternlicht, 2005. Key stages in mammary gland development: The cues that regulate ductal branching morphogenesis. Breast Cancer Res 8(1):201.

Strand, 2012. A proteomic analysis of differentiation in the mammary epithelium. Unpublished thesis.

Thompson, et al., (eds.), 2009. *Milk Proteins: From Expression to Food*. New York: Academic Press.

Toogood, et al., 2007. Dynamics driving function: new insights from electron transferring flavoproteins and partner complexes. FEBS J 274(21):5481-5504.

Treiber, et al., 2012. Regulation of microRNA biogenesis and function. Thromb Haemost 107(4):605-610.

Trott, et al., 2012. Triennial lactation symposium: Prolactin: The multifaceted potentiator of mammary growth and function. J Anim Sci 90:1674-1686.

Turkington, et al., 1968. The hormonal control of lactose synthetase in the developing mouse mammary gland. J Biol Chem 243:3382-3387.

Vakharia and Mizejewski, 2000. Human alpha-fetoprotein peptides bind estrogen receptor and estradiol and suppress breast cancer. Breast Cancer Res Treat 63:41-52.

Vernon, 1989. Endocrine control of metabolic adaptation during lactation. Proc Nutr Soc 48:23-32.

Wainwright, 1931. Comparison of conditions associated with breast cancer in Great Britain and America. Am J Cancer 15:2610-2645.

Walden, et al., 1998. Evidence that the mammary fat pad mediates the action of growth hormone in mammary gland development. Endocrinology 139(2):659-662.

Wang and Oster, 1998. Energy transduction in the F1 motor of ATP synthase. Nature 396(6708):279-282.

Waseem, et al., 1999. Keratin 15 expression in stratified epithelia: Downregulation in activated keratinocytes. J Invest Dermatol 112:362–369.

Watson, 2006. Involution: Apoptosis and tissue remodeling that convert the mammary gland from milk factory to a quiescent organ. Breast Cancer Res 8(2):203.

Weisz-Carrington, et al., 1977. Plasma cells and epithelial immunoglobulins in the mouse mammary gland during pregnancy and lactation. J Immunol 119(4):1306-1309.

Welsh, et al., 2003. Vitamin D-3 receptor as a target for breast cancer prevention. J Nutr 133:2425S–33S.

Wiesen et al., 1999. Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development. Development 126:335-344.

Williams and Daniel, 1983. Mammary ductal elongation: Differentiation of myoepithelium and basal lamina during branching morphogenesis. Dev Biol 97:274-290.

Wolkenhauer and Hofmeyr, 2013. Interdisciplinarity as both necessity and hurdle for progress in the life sciences. In Andersen, et al., (eds.), *New Challenges to Philosophy of Science - The Philosophy of Science in a European perspective 4*. Amsterdam, The Netherlands: Springer Verlag.

Wood, et al., 2000. The insulin-like growth factors (IGFs) and IGF binding proteins in postnatal development of murine mammary glands. J Mammary Gland Biol Neoplasia 5:31–42.

Yang, et al., 2009. Regulation of β4-integrin expression by epigenetic modifications in the mammary gland and during the epithelial-to-mesenchymal transition. J Cell Sci 122(14): 2473–2480.

Yip and Knox, 1972. Function of arginase in lactating mammary gland. Biochem J 127:893-899.

Yoshida, et al., 2013. Molecular chaperone TRAP1 regulates a metabolic switch between mitochondrial respiration and aerobic glycolysis. Proc Natl Acad Sci U S A 110(17):E1604-E1612.

Zinger, et al., 2003. Prolactin expression and secretion by human breast glandular and adipose tissue explants. J Clin Endocrinol Metab 88(2):689-696.

Zinser, et al., 2002. Vitamin D(3) receptor ablation alters mammary gland morphogenesis. Development 129:3067–3076.

APPENDIX A – Use Case Document from CSC 366

CSC 366
Winter 2012
Course Project

Regulatory Sequences Database

Project Documentation
Use Cases

Version 2.0

Dan Peterson
Alyssa Conly
Laura Strand

Alex Dekhtyar

1. General Information

This document outlines the use cases for the database of regulatory sequences found in the promoter regions of up- or down-regulated genes discovered when comparing two states of a cell to each other.

In what follows, the term *database* is used to refer to ***both*** the collection of data described in the companion document, ***and*** to the software system you are being asked to design and develop in this course. Where necessary, we disambiguate by referring to the latter as the *database software* or the *database application*.

There are two ways to implement the database software: as a stand-alone application that connects to a local database (or a database on a local network) for data, and as a web-based application that runs inside the browser. Except for some user management issues, discussed at the end of this document, all other use cases do not differentiate between the two approaches. The database software is to be designed as essentially a single-user system. The necessary on-line precautions are discussed at the end of the document.

The database functionality is separated into three broad categories:

- Data insertion/manipulation functionality;
- Data browsing/filtering functionality;
- Analytical fnctionality.

Data insertion functionality consists of use cases for inserting the data into the database, as well as the functionality for modifying the data that is already in the database.

Data browsing and filtering functionality consists of use cases for showing the users the contents of the database in a convenient form.

Analytical functionality consists of use cases the users of the database plan to use in their research pursuits.

2. Data Insertion and Manipulation Functionality

[UC-1.] Bulk Insertion/Insertion of data for a single experiment. Most of the time, when the database is updated in a significant way, the events immediately preceding the update involve a researcher/database user performing an *in-vitro* experiment comparing two cell states, determining the proteins that were *different in abundance*, running TESS queries for **all discovered proteins/genes** and collecting TESS answers in a group of spreadsheets residing in a single directory.

Because of this, the core database update/data insertion functionality is the **bulk insertion** of the data for one *experiment* or more experiments (if data for more than one experiment is available at the same time).

The data to the database will be provided in a collection of CSV files representing individual worksheets of the Excel spreadsheets in which TESS returns the data.

Information provided. The user selects the bulk update functionality from the main menu/main screen of the database software. The software provides the *bulk insertion* dialog. The user selects a local directory that contains the CSV files for the experiment. This action is subject to the following restrictions and specifications:

- The directory will contain four CSV files for each discovered in the experiment up- or down-regulated gene. The four files storing information about the same gene will have similar filenames, different only by the designation of the worksheet from the Excel file, which is used. The four types of CSV files are:
  1. *Experiment description (Job parameters).* This CSV file will contain <Name, Value> pairs describing the experiment and the gene.
  2. *Sequence.* This CSV file will contain the 2000 character promoter region sequence for the gene.
  3. *Regulatory sequence discoveries.* This CSV file will contain information about discoveries of individual regulatory sequences in the promoter region of the gene.

- The filename conventions are specified in the **Lab 5** assignment.
- In the past CSC 366 course we discovered that some software development frameworks used by students in the course lacked the predefined facilities for selection of a *single directory* (rather than selection of a file, or multiple files). Please know that ***it is highly desirable for the software for the users to be able to select directories, not files.***

Software actions. Upon selection of the directory, the software shall scan the CSV files in the directory. The software shall create a list of gene names, the information about which is contained in the CSV files. For a gene name to be included on the list, *all four* CSV files for this gene must be present in the directory. The software then, displays the list of genes to the user. The user inspects the list of genes, and is provided with the opportunity to exclude individual genes from the upload. Upon completion of the gene selection, the user presses the ***Upload*** button. The software reads, one by one, the files for each gene, and updates the database with the information on the experiment, the gene, its promoter region, the regulatory sequences found in the promoter region and the

transcription factor summaries. The software may provide the user with some feedback concerning the ongoing insertion activities, as the bulk upload proceeds.

Note. This use case applies to two situations: a *bulk upload proper* used to populate an empty database, and an *addition of a new experiment* to the database. In the former case, the database is empty, and the data to be inserted contains, potentially, multiple experiments. In the latter case, the data for a single experiment becomes available and is uploaded in bulk. The system behavior should be the same in both cases, except for «noticing» how many different experiments are being uploaded.

[UC-2.] Loading information about a single gene. To make it easy for the users to compensate for errors of omission (should any Excel spreadsheets from which the CSV files are constructed become corrupt), this use cases offers the users of the system the opportunity to upload information about a single gene found up- or down-regulated in a *specific experiment*. The uploaded data is available in four CSV files.

Information provided. The user selects «Upload Single Gene Data» functionality from the main menu/screen of the software. The system brings up the «Upload Single Gene Data» dialog screen. On the screen, the user navigates to a directory that contains the files for upload. When the directory is selected, the software reads the list of CSV files in the directory and determines the names/labels of all genes, whose complete descriptions are found. A description is complete if all four CSV files describing the data to be stored in the database are found. The software outputs the list of available experiments and gene data to the user. The user selects the gene name the data for which (s)he wants to upload and presses the «Upload» button (or initiates the upload using some other interaction with the dialog screen)

Software Actions. The software reads the contents of the four CSV files associated with the gene selected by the user, verifies the information about the cell comparison and the experiment (and if they are not represented in the database yet, inserts this information) and uploads the information about the gene, its promoter region, discovered regulatory sequences and transcription factor summaries to the database.

[UC-3.] Editing the contents of the database. The software shall provide the facilities for updating individual records stored in the database. In Section 3 of this document, the browsing functionality is detailed. As part of the browsing functionality the user shall be given the ability to view individual records. It is convenient to incorporate the *update record* functionality and *delete record* functionality together with the functionality for viewing records.

2. Browsing and Filtering Functionality

[UC-4.] Hierarchical browsing of the database. This functionality can be implemented on either the front page of the database application, or on a page that is easy to access from the front page.

The key hierarchical organization of the data in the database (as described in the prior documentation) is as follows:

- The researchers are interested in *comparing two different states* of the same cell from a specific specie.
- For each such *comparison*, researchers can conduct multiple *experiments*.
- With each experiment, a collection of *genes* found to be different in abundance in the two cell states is associated.
- For each *gene* identified this way, a list of *regulatory sequences* and associated *factors* found in its promoter region is stored.

One of the main mechanisms for study of the data by the users revolves around browsing the data in this hierarchical manner.

User-software interaction.

- The user selects to pursue the hierarchical browsing functionality of the software (if the hierarchical browsing is enabled on the front page of the software, then what follows occurs without the user making an explicit selection of activity).
- The software displays a list (or a table) of species, currently available in the database.
- The user selects the specie of interest.
- The software, *while keeping the list of species on-screen,* produces a list of comparisons, currently available in the database. For each comparison, include the name of the *specie* for which it is made.
- The user selects the comparison of interest.
- The software, *while keeping the list of the comparisons on-screen*, produces a list of available *experiments* for the given comparison. The information displayed within "Experiment" includes date of the experiment and the name of the person who staged it, the number of genes identified, the number of genes showing "Regulation" up or down.
- The user selects the experiment of interest.
- The software, *while keeping the list of species, comparisons and the list of experiments on-screen,* produces the list of all genes found different in abundance for the experiment. For each gene, its name, location (chromosome, begin site, end site), whether it is up- or down-regulated, and the total number of discovered factors.
- The user selects a gene of interest.
- The software, *while keeping all previous lists on-screen* produces the list of all factors found in the promoter region of the gene. For each factor, its origin (model), the name of the factor, the number of times it is found in the promoter region.
- The user selects a factor of interest.
- The software, *while keeping all previous lists on-screen* produces the list of all regulatory sequences found for that factor in the promoter region of the gene. For each regulatory sequence, its position (beginning, length, sense), each "quality" measure (all the "*l-values*") and the actual string shall be displayed.

- The user selects a regulatory sequence
- The software shows the user a dialog screen for viewing (as well as editing) information about a single regulatory sequence (see use case **[UC-8]**).

[UC-4.1] User filtering of the displayed information. While the user is perusing hierarchical browsing, the following filtering options shall be made available.

- For the ***list of genes*** for a given experiment, the software shall provide the user with the ability to select *all genes, up-regulated genes only* or *down-regulated genes* only.
- For the ***list of regulatory elements*** for a given gene (in a given experiment), the software shall provide the ability to
    - Sort the list by any displayed column
    - Filter the list by any of the ***l-values*** stored in the database. The user provides a range of values, or a lower bound on the ***specific l-value parameter*** (any single ***l-value parameter*****).** The software keeps information about regulatory elements whose ***l-value*** is in the appropriate range.
    - Location on the promoter sequence. The user provides information about the range ("from" and "to" positions) for the *beginning of the regulatory element* and, optionally, select the sense of the regulatory element. The software keeps in the list only the regulatory elements with starting points in the specified range (and with the specified sense, if one is provided).

[UC-5.] Summary list. [DEPRECATD]

[UC-6.] Database Summary: Genes.

- The user selects "Database Summary: Gene list" from the main menu/main screen of the system.
- The software brings up a dialog screen and populates it with the list of genes found in the database. For each gene, the following information should be provided:
    - Species (i.e., the same gene found in different species should show up multiple times in the table)
    - Gene name
    - Gene abbreviation
    - Gene location (chromosome, begin site, end site)
    - Number of comparisons in which the gene was found different in abundance (all occurrences, down regulated occurrences, up regulated occurrences)
    - Number of experiments in which the gene was found different in abundance (all occurrences, down regulated occurrences, up regulated occurrences)

- The user can sort the genes by any column in the output.
- The user selects a gene of interest.
- The software opens a dialog screen containing the information about an individual gene. The new screen contains all information about the gene available from the database (name, location, etc), as well as the list of experiments in which the gene was found to be different in abundance as well as the regulation (up or down) for each experiment.

[UC-7.] Database Summary. Regulatory elements/Transcription factors.

- The user selects "Database Summary: Transcription factors" functionality.
- The software brings up a dialog screen and populates it with the list of all transcription factor names found in the database. For each transcription factor, the list contains the following information:
  - Transcription factor name
  - "FAC" (if there is one present)
  - Number of models in which this transcription factor was found anywhere in the genes of interest (i.e., those that were different in abundance in any of the stored experiments)
  - Number of genes in which this transcription factor was found
  - Total number of occurrences of the transcription factor in the database
- The user can perform the following manipulations with this list:
  - Select a transcription factor of interest. In response, the software brings up a dialog screen populated with the list of all occurrences of the given transcription factor in the regulatory elements in the database. The list can be sorted by each column it contains (comparison name, experiment information, gene name, regulatory element location, model number).
  - Sort the list of transcription factors by any column in the list, in particular by the number of models, the number of genes and the number of occurrences.

[UC-8.] Viewing information about an individual regulatory element for a gene.

At the bottom of many navigational hierarchies in the system lies the dialog screen that shows the information on a single regulatory element for a given gene from a given experiment. The dialog screen showing this information shall contain the following:

- The information about the comparison and the name of the experiment.
- The information about the gene for which the regulatory sequence is found.
- The information about the regulatory element itself:
  - Its location in the promoter sequence (beginning, sense and length);
  - The nucleotide sequence;
  - The origin of the regulatory element discovery (model);
  - The quality of the match (all the *l-values*).
  - The name of the transcription factor(s) associated with the sequence

Additionally, the dialog screen shall provide for the following types of interaction with the user:

- Editing the information about the regulatory element.
- Deleting the information about the regulatory element.
- Searching for *similarly described/duplicate* regulatory elements in the given promoter sequence. Searching for this information shall result in a list of *similar* regulatory elements displayed on the dialog screen.
  A regulatory sequence is *similar* to another regulatory sequence in the same promoter region if both sequences start at the same position, go in the same direction and their lengths are within 1 nucleotide of each other.

3. Analytical Functionality

Note: UC-4.1, User filtering function would be very helpful to apply to all the UCs within the analytical functionality. I.e. allow the user to filter all output according to the specifications in UC-4.1. Part of this has been added as specific components of UC-9 and UC-10.

[UC-9.] Finding all genes that have specific transcription factor or factors present.

One of the key purposes of this database is to allow the researchers to study similarities between different genes found to be important for a comparison of two states of a cell. One of the key analytical tasks is to find and report a list of genes that contain a specified single transcription factor or all specified transcription factors from a given list.

Inputs provided by the user. The user supplies the following information to the software:

- The transcription factor or factors of interest.
- Constraints for quality parameters for factors (*L-values*) that, if provided by the user, limits the analysis to factors that conform to the constraints.
  - Minimum L a
  - Minimum L a/
  - Minimum L q
  - Maximum L d
- The scope of the search. The scope of the search can be an entire database, a single specie, a single comparison, or a single experiment.

The software **shall provide convenient, user-friendly UI** for supplying the input information.

Information provided by the software. Upon accepting the inputs provided by the user the software searches the specified part of the database for genes, whose lists of discovered regulatory elements contain elements to which **all** transcription factors from the input list bind. For each gene, the software shall return the following information:

- Gene name
- Comparisons in which the gene was identified and Regulation (UP or DOWN) for each comparison
- Experiment for each comparison

The software shall provide the opportunity to reorder the output by the contents of any column.

[UC-10.] Listing transcription factors by popularity in the database.

Another key purpose of this database is to allow the researchers to identify factors that are centrally involved in the comparisons being studied. This can be accomplished by allowing the researchers to select multiple comparisons (or all comparisons) and listing all of the factors that appear in any of the genes identified from those comparisons, sorted by number of appearances.

Inputs provided by the user. The user supplies the following information to the software:

- Species (can be one or multiple)

- For each species selected, a list of all available comparisons appears, and user selects one or more (up to all) comparisons.
- For each comparison, if there is more than one experiment, the user can select one or more experiment.
- Constraints for quality parameters for factors (*L-values*) that, if provided by the user, limits the analysis to factors that conform to the constraints.
    - Minimum L a
    - Minimum L a/
    - Minimum L q
    - Maximum L d

The software **shall provide convenient, user-friendly UI** for supplying the input information.

Information provided by the software. Upon accepting the inputs provided by the user the software searches the specified part of the database for factors that appeared in any gene included in the user inputs, and provides a list of all such factors, ordered by number of times the factor appeared in the portion of the database included in the search. The resulting list includes:

- Factor (starting with most "popular" or highest number of appearances)
- Number of times the factor appeared
- Number of genes in which that factor appeared
- When the factor is chosen, list of genes in which the factor was identified
- For each gene, comparison(s) and experiment(s) in which the factor was identified

[UC-11.] Listing factors found in specific regions of the promoter. [REMOVED]

[UC-12.] Listing genes, which have transcription factors in specific regions. [REMOVED]

[UC-13.] Finding genes that share multiple transcription factors [EXTRA CREDIT].

Explanation. This is an **EXTRA CREDIT** use case. It involves implementation of the Apriori algorithm (or any other algorithm for computing the list of frequent itemsets).

The nature of the task is as follows. With each gene stored in the database, a list of transcription factors found to bind to its promoter region is associated. It is of interest to the users of the system to be able to find the lists of transcription factors that are commonly shared, and be able to view the list of genes that share these transcription factors.

An *itemset* as applied to this task, is a collection of transcription factor names (associated with a single gene). Given a list of genes, the ***support of an itemset of transcription factors*** is the percentage of the genes, which contain the given itemset of factors.

An itemset is *frequent* if its support is above some value $0<=a<=1$, specified by the user. The goal of this use case is for the software to determine the list of frequent itemsets in your database, and to provide the list of applicable genes for each frequent itemset.

An *itemset* is called a *skyline itemset* if it is *frequent* by has no frequent supersets. The software should report only *skyline itemsets* (lest, there may be way too many itemsets to report).

There is a number of algorithms for discovery of frequent itemsets, most of the relying on the ***Apriori principle***:

Any subset of a frequent itemset must be frequent.

Inputs provided by the user. The user will provide the software with the following information:

- The filter on the list of genes to be searched (all genes, genes from a specific experiment or a collection of experiments)
- The *frequency threshold a* : a number between 0 and 1, representing the percentage of all genes in which an itemset must appear to be reported, ***or*** a *minimum number of occurrences threshold*: a number of genes in which the itemset is to appear to be reported.
- Itemset size thershold: the smallest possible size of a reported frequent itemset.

Information provided by the software. Given the parameters above, the software shall implement the Apriori algorithm (see below), or any modification of the Apriori algorithm deemed sensible by the development team to find the list of all *skyline frequent itemsets* that match the frequency threshold and itemset size threshold.

The frequent itemsets shall be reported in a convenient succinct manner. Selection of a frequent itemset shall lead to a dialog window which shows the list of all genes that have all the transcription factors from the selected itemset associated with them.

[UC-14.] Sort lists of transcription factors by "quality". [REMOVED]

[UC-15.] Promoter Region Map [EXTRA CREDIT].

This is an **EXTRA CREDIT** use case. Bioinformatics databases that store genomics data often display information about various places of interest (genes, regulatory sequences, etc...) using a *genomic map*. A *genomic map* in this context, is a representation of a *linear DNA sequence* using UI or graphical elements, that shows, via color-coding, and/or other means, where on the observable part of the sequence various "interesting" places lie. The map is usually responsive to user interaction: e.g., a user can click a mouse button on a highlighted location to receive more information about it.

For the software system you are building, a map can be created for the promoter region sequences. A promoter region is associated with a single gene of a specific species, so, a promoter region map can be displayed and used for interactive purposes whenever information about regulatory sequences

of a single gene (from a specific experiment) is shown. Additionally, there may be value in showing multiple promoter region maps together, at it might provide some visual cues for the software users about the similarities between regulatory element occurrences in different promoter regions.

A promoter region map is essentially a color-coded, and/or otherwise marked up *line* (represented as a graphical element and/or in some other way), in which an internal mapping between each physical (pixel) position on the line and a location in the promoter region is established. (You can also represent the map as multiple lines: e.g., one for each DNA strand). The map shall, at least, contain references (color-coded regions, tick marks, or some other ways) to the location of different regulatory elements in the promoter region. The references shall be "clickable" in a sense that there should be a way to interact with the promoter region map and "select" a specific mark/region. Selection of such a region shall result in additional information (essentially a record, or records about the regulatory sequences associated with the selected location) becoming visible/available/highlighted/selected in the current dialog window.

Even a simple promoter region map may be of great use to the users of the system.