DYNAMIC ALTERNATIVE POLYADENYLATION (APA) EVENTS IN MAMMARY GLAND DEVELOPMENT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÜLTEN TUNCEL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

JUNE 2016

Approval of the thesis:

DYNAMIC ALTERNATIVE POLYADENYLATION (APA) EVENTS IN MAMMARY GLAND DEVELOPMENT

Submitted by **GÜLTEN TUNCEL** in partial fulfilment of the requirements for the degree of **Master of Science in Department of Biology, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Orhan Adalı _____
Head of Department, **Biology**

Assoc. Prof. Dr. A. Elif Erson-Bensan _____
Supervisor, **Biology Dept., METU**

Examining Comitee Members:

Prof. Dr. Mesut Muyan _____
Biology Dept., METU

Assoc. Prof. Dr. A. Elif Erson Bensan _____
Biology Dept., METU

Assist. Prof. Dr. Nihal Terzi Çizmecioğlu _____
Biology Dept., METU

Assoc. Prof. Dr. Tolga Can _____
Computer Eng. Dept., METU

Assist. Prof. Dr. Özgür Şahin _____
Molecular Biology and Genetics Dept., Bilkent Univetsity

Date: 07.06.2016

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Gülten Tuncel

Signature :

ABSTRACT

DYNAMIC ALTERNATIVE POLYADENYLATION (APA) EVENTS IN MAMMARY GLAND DEVELOPMENT

Tuncel, Gülten

M. S., Department of Biology
Supervisor: Assoc. Prof. Dr. A. Elif Erson Bensan

June 2016, 74 pages

Mammary gland is a dynamic organ that undergoes dramatic changes in response to age, circulating hormones and reproductive status. The extensive growth, regeneration and remodelling abilities of the mammary gland is of specific interest to better understand the molecular mechanisms deregulated in breast cancers. To this end, we hypothesized alternative polyadenylation (APA) as a mechanism that generates mRNA isoforms with different lengths, to be dynamically regulated at different stages of mammary development. APA is known to be regulated in a tissue and developmental stage specific manner and is linked to mRNA 3'UTR isoform variability, which may result in altered protein levels and/or functions. We used a combinatorial *in silico* and *in vitro* approach and delineated a dynamic pattern of APA in gestation, lactation and involution stages. Experimental confirmation of 3'UTR isoforms indeed provided preliminary data for future studies to provide a possible link between mammary tissue development, APA and breast cancer.

Keywords: Alternative Polyadenylation, Mammary Gland Development, Breast Cancer

ÖZ

MEME DOKUSU GELİŞİMİ SIRASINDA DİNAMİK ALTERNATİF POLİADENİLASYON (APA) OLAYLARI

Tuncel, Gülten

Yüksek Lisans, Biyoloji Bölümü
Tez Yöneticisi: Doç. Dr. A. Elif Erson Bensan

Haziran 2016, 74 sayfa

Meme dokusu bireyin yaş, hormon seviyelerindeki değişim ve üreme durumuna göre sürekli değişim halinde olan dinamik bir organdır. Meme dokusunun kapsamlı büyüme, yenilenme ve yeniden şekillenme yetisi, meme kanserinde deregüle olan moleküler mekanizmaları daha iyi anlayabilmek için özel ilgi görmektedir. Bu durumda, farkli uzunluklarda mesajcı RNA (mRNA) izoformlarının sentezlenmesine yol açan alternatif poliadenilasyon (APA) mekanizmasının meme dokusu gelişimi sırasında farklı şkillerde regüle olabileceği hipotezini öne sürdük. 3’ UTR (transle olmayan bölge) izoformlarında varyasyona sebep olarak değişmiş protein seviyeleri ve/veya fonksiyonlarına neden olabilen APA’nın farklı doku ve gelişim aşamalarına özel regüle edildiği bilinmektedir. Biz bu çalışmada *in silico* ve *in vitro* yaklaşımları kombinatoryal şekilde kullanarak meme gelişiminin hamilelik, emzirme ve geri çekilme süreçlerinde dinamik APA olaylarını tanımladık. Değişen 3’UTR izoformlarını doğrulayan deneysel çalışma sonuçlarımız da gerçekten meme dokusu gelişimi, APA ve meme kanseri arasındaki olası bağlantıyı gösterebilecek gelecek çalışmalar için ön veri sağlamıştır.

Anahtar Kelimeler: Alternatif Poliadenilasyon, Meme Dokusu Gelişimi, Meme Kanseri

To my family

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my thesis advisor Assoc. Prof. Dr. A. Elif Erson Bensan for her encouragement, endless support and immense knowledge. The door to her office was always open whenever I needed advice about my research.

I would like to thank my thesis committee members; Prof. Dr. Mesut Muyan, Assist. Prof. Dr. Özgür Şahin, Assoc. Prof. Dr. Tolga Can and Assist. Prof. Dr. Nihal Terzi Çizmecioğlu.

I am grateful to all members of the Erson Lab; Merve Öyken, Esra Yavuz, Tuna Çınkıllı, Ayça Çırçır Hatıl, Begüm Akman Tuncer, Oğuzhan Beğik and Murat Erdem for being amazing friends, helping me adjust to a new country very quickly and of course for sharing their knowledge. I would also like to thank all Muyan Lab members for their friendship and for sharing their resources whenever I needed.

I am always particularly thankful to Mine Hüryaşar, who has been my spiritual sister since primary school years, for helping me stay sane through my undergraduate and postgraduate studies overseas.

I must express my very profound gratitude to my fiancé Mustafa Dereboylu, who has always been by my side with his love to motivate and support me through all ups and downs of my journey.

Finally, I wish to offer my deepest thanks to my parents Şaziye Tuncel and Şeref Tuncel as none of this would have been possible without their endless love, support, trust and patience by all means. I also warmly appreciate the support and understanding of my extended family.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1. Mammary Gland Development

The mammary gland is a very complex organ, which undergoes extensive remodelling in shape, size, structure and function starting from intrauterine life up until menopause. These changes are in association with the individuals age, circulating hormone levels, as well as the reproductive status (Figure 1.1) [1]. It is in general the endocrine system that coordinates the development of the mammary gland [2].

The mammary gland parenchyma is known to be arising from a single epithelial ectodermal bud. The mammary-specific progenitor cells found in the primitive nipple epithelium start to proliferate as early as 5-6 weeks of gestation in humans [3]. Even though the final morphology of the breast at birth is still not clear, it is thought that the progenitors proliferate to form ducts ending in short ductules or lobes lined by one or two layer of luminal epithelial and one layer of myoepithelial cells [1]. These structures altogether compose the primitive mammary gland structures in the newborn breast (Figure 1.1). Breast is the only organ in the body that is not fully developed until a full-term pregnancy in women [4].

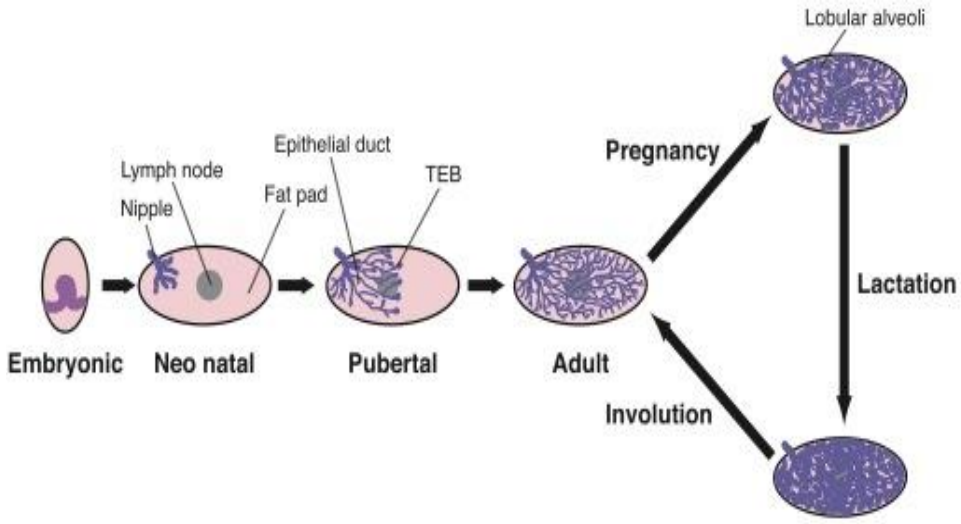

Figure 1.1. Stages of normal mammary gland development.
The mammary gland is specified at early embryonic stages in all mammals. The epithelium invades through the fat pads and forms a small, branched ductal network. After birth, mammary gland grows in concert with the infant until puberty. Once the ovarian hormones start to be released in puberty, terminal end buds (TEBs) are formed and the ducts invade further through the fat pads. Whole mammary gland is filled with ductal branches in an adult individual. During pregnancy, the epithelial cells restart to proliferate, forming side branches and alveolar structures on the expanded ductal tree with secretory function. The lobular alveoli are stimulated primarily by prolactin and oxytocin to provide milk for suckling infants upon parturition. During involution, the epithelial tissue undergoes major apoptosis, the small adipocytes redifferentiate back to larger forms and the gland is remodelled back to a state similar to an adult individual. Figure taken from [5].

Until the approach of puberty, the mammary gland development is mostly limited to the general growth of the body. However, when the infant reaches puberty, ovulation and establishment of regular menstrual cycles trigger massive morphological and histological changes in the breast [6]. The overall increase in breast size in this period is primarily due to the increased deposition of adipose tissue within the mammary gland [7]. The distinct changes in epithelial and stromal layers, on the other hand, are fueled by an ovarian hormonal circuit, which acts on mammary stem cell (MaSC) populations [2]. Cap cells found at the basal surface of terminal end buds (TEBs) lead the expansion of the ducts into fat pads. During extension, bulk of epithelial body cells undergoes apoptosis to allow the formation of hollow ductal structures (Figure. 1.2.) [8]. These changes result in the elongation of the existing primitive ducts into branches that terminate to form clusters called lobules (Figure 1.1) [1]. This lobular structure of

the mammary gland composes the adult virgin mammary gland structure and continues to develop gradually during adolescence through to adult life [4].

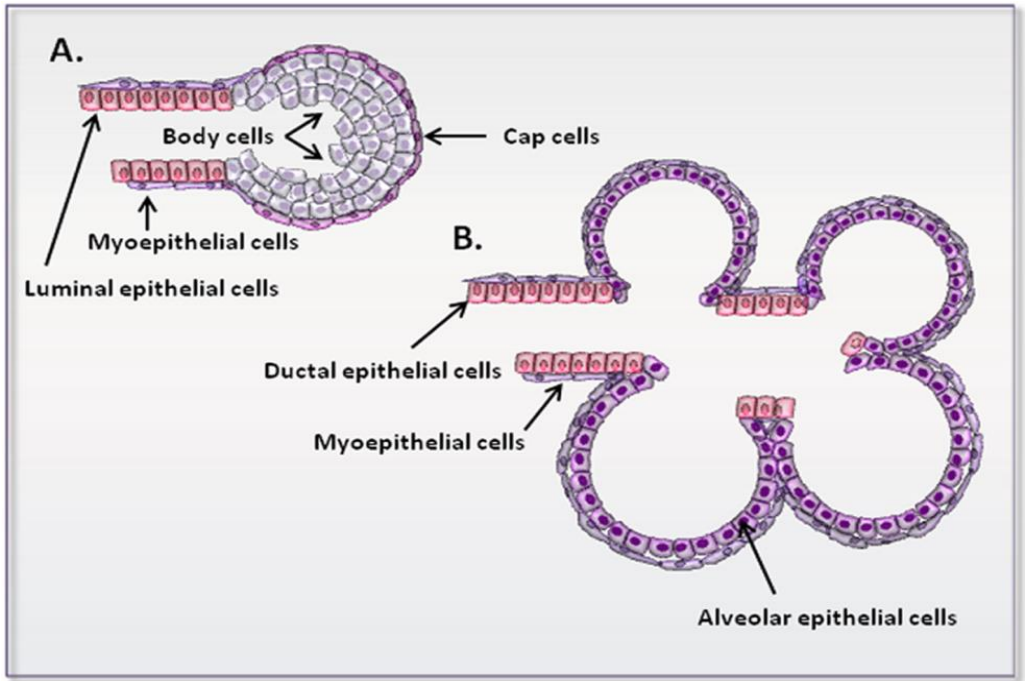

Figure 1.2. Representation of cellular structures within mammary gland on different developmental stages.
(A) The ductal structures in prepubertal and pubertal mammary gland are composed of luminal epithelial cells covered by a single layer of myoepithelial cells and terminate by terminal end bud (TEB) structures, which have high proliferative capacity. TEBs have multilayered body cells that are surrounded by the cap cells. Caps cells have contact with the basal lamina Elongation of the ducts is lead by the cap cells. Bulk of body cells undergo apoptosis and result in single layer luminal epithelial cells. (B) During pregnancy, by the action of circulating hormones, extensive branching of the ducts and formation of alveolar structures takes place. The alveolar structures contain one layer of luminal epithelial cells, capable of producing milk during lactation (lactocytes), surrounded by a layer of myoepithelial cells, and the basement membrane. When stimulated by oxytocin, myoepithelial cells contract and the milk produced within the lactocytes is released into the ducts and further to the nipple. Figure is taken from [9].

The complete remodeling of the breast happens when it undergoes a pregnancy-lactation cycle (PLC) as it completely transforms into a fully mature, milk-giving organ. Elevated levels of the circulating lactogenic hormone complex, which includes estrogen, progesterone and prolactin regulates maturation process [10]. Among these, there are several other hormones and growth factors that directly regulate mammary gland expansion during pregnancy such as the placental lactogen, oxytocin, cortisol, transforming growth factor alpha (TGFα) and epidermal growth factor (EGF),

whereas other factors such as growth hormone, insulin, leptin, fibroblast growth factors and glucocorticoids exert their effects indirectly [11]. By the cumulative action of these factors during the first months of pregnancy, MaSCs undergo an initial phase of proliferation that result in the formation of new ductal structures, elongation of the former ducts via mitosis at the TEBs and formation of spherical structures, named as alveoli, at the terminal sites [12]. Newly forming ducts invade through the fat pads in the mammary gland.

In the second trimester of pregnancy, following the expansion phase, cellular differentiation process at the alveoli and luminal epithelial cell layer were triggered by increasing levels of prolactin hormone. This results in the formation of lactocytes. Lactocytes are the milk-producing cells in the mammary epithelium, which synthesize and secrete the main milk proteins such as beta-casein in response to prolactin stimulation. This period is accompanied by the accumulation of colostrum (first secretion) inside the ducts and alveoli [6]. However, high levels of circulating progesterone and estrogen blocks the prolactin's stimulatory effect on milk synthesis [13].

After 30-40 hours of parturition, circulating levels of progesterone decrease rapidly as the placenta is delivered during birth. Together with increased levels of prolactin and the stimulation of oxytocin release by infant suckling, the myoepithelial cells contract resulting in the expulsion of milk into the ductal lumen. Milk is then ejected from the nipple [6]. This continues until the mother stops breastfeeding the infant.

At the end of breastfeeding, the bulk of mammary epithelium undergoes significant apoptosis and the tissue restructures itself to a state resembling that of a pre-pregnant individual. This stage of breast remodeling is called involution (Figure 1.1). During this period, a bulk of epithelial cell death, stromal tissue remodeling and immune cell activation is observed, similar to microenvironments observed in wound healing and also tumor progression [14]. These changes are followed by the decreased expression of genes involved in milk protein synthesis to basal levels and also the genes

associated with regulation of differentiation and proliferation, p53, TGF-β1 and c-myc [15]. The hormones involved in the differentiation of mammary tissue and their action mechanisms will be discussed in the next sections.

1.2. Breast Cancer Risk Related To Pregnancy-Lactation (P-L) Cycle

Numerous epidemiologic studies have reported a possible link between pregnancy, breastfeeding and breast cancer risk. In general, a complete P-L cycle and especially breastfeeding the infant is hypothesized to protect the breast against neoplastic transformation [16]. There are two possible hypotheses to explain this observation. 1) A full-term pregnancy and lactation period reduces the number of menstrual cycles a woman experiences and thus reduces her cumulative exposure to endogenous hormones. 2) The expanded population of epithelial progenitor cells during pregnancy are lost progressively during the lactation period through terminal differentiation and overall proliferative activity of the mammary epithelium is decreased through this differentiation process [17].

A recent study have demonstrated that women with basal-like tumours are less likely to have breastfed before breast cancer diagnosis, regardless of duration, than woman with luminal A tumours at diagnosis. Also, a 30% decreased risk of tumour recurrence is observed in women that breastfed, and breastfeeding for at least six months or more conferred a slightly larger reduction in recurrence risk. However, among HER-enriched tumours the association were suggestive of increased risk [17].

Another group has experimentally shown that the first pregnancy results in the expression of a specific genomic signature in the breast tissue, resulting from the completion of the organ's differentiation cycle. Furthermore, the genetic signature formed in this stage is thought to be a biomarker linked with a possible decrease in breast cancer risk [16].

However, a clear mechanistic understanding behind the association between breast cancer risk and P-L cycle is currently missing, partly because most epidemiological studies did not distinguish breast cancer subtypes. Breast cancer is recognised to harbour three molecular subtypes that are; luminal like (ER+), ERBB2+ and basal like (ER-, PR-, HER2- or TNBC), suggesting the underlying pathologies to be different [18]. Therefore it will be of great interest to delineate the increased or decreased risk associated with P-L cycle in different breast cancer subtypes. In addition, most epidemiological studies compare the occurrence of breast cancer events in nulliparous women versus women that gave birth and breastfed at least once in their lifetime. Therefore understanding the association between pregnancy and breast cancer and breastfeeding and breast cancer is challenging.

In summary, most breast tumours originate from the undifferentiated precursor cells in the epithelia or ductal lobes. Because of their highly proliferative nature during the gestation phase, carcinogen binding risk to the DNA and low reparative capability the undifferentiated cells within the breast have increased risk of undergoing neoplastic transformation. Therefore, lactation phase of the P-L cycle may reduce this susceptibility by inducing terminal differentiation.

However, on the other side of the medallion it is clear that there are many parallels between P-L cycle and breast tumorigenesis. Normal developmental processes such as invasion through the fat pads while branching, re-initiation of cell proliferation in the terminal end bud cells during gestation and resistance to apoptosis especially in the lactating mammary gland to protect the gland from premature involution and also the ability to induce angiogenesis are processes that are highly relevant to breast tumorigenesis [5, 19].

The interaction between epithelium and extracellular matrix (ECM) is also very critical and must be strictly controlled during branch morphogenesis because the fate of progenitor cells within the mammary epithelia is determined by the factors within the ECM [20, 21]. Therefore, uncontrolled expansion of the progenitor cell population

during gestation is thought to increase the risk of possible malignant transformation cases in the breast epithelia [22-24].

Genes that are found to be actively upregulated during branching have also been implicated in breast cancer. These include matrix matallopeptidases (MMPs) such as MMP3 and MMP14, several growth factors and components involved in protein kinase pathways such as wingless-type MMTV integration site family, member 1 (WNT1) [5].

Protein kinase receptors and their ligands play key roles during mammary gland morphogenesis and aberrations in their expression or function of these proteins and their upstream or downstream signalling components can easily initiate malignant transformation of the breast [25, 26]. Four major pathways that are important in P-L cycle and also known to be altered in breast tumorigenesis are discussed below.

First is the prolactin induced signalling. The cytokine prolactin (PRL) is an important molecule that has a wide range of physiological functions. It plays the key role in growth and differentiation of mammary gland into its mature form during pregnancy and initiation of milk protein synthesis in lactation. It was shown to stimulate cell proliferation during pregnancy-lactation cycle in mammals and has been implicated to play a role in tumorigenesis [2, 27]. Additionally, PRL also has a role in regulating the immune response, parental behavior and shown to stimulate neurogenesis in maternal and fetal brains [28-30]. There are various variants of the prolactin protein, each resulting from post-translational modifications of the major monomeric 23kDa isoform such as phosphorylation, glycosylation, deamination and sulphation. Additionally 14kDa, 16kDa and 22kDa prolactin variants are generated from proteolytic cleavage of the 23kDa protein [31].

Prolactin is synthesized and secreted primarily from the lactotroph cells of the anterior pituitary gland but it can also be secreted from extrapituitary sites including myometrium, breast, lymphocytes and leukocytes in humans [6, 32]. Synthesis and

secretion is controlled very strictly, involving prolactin stimulatory as well as inhibitory factors. Pituitary prolactin synthesis is regulated by the Pit-1 transcription factor, which binds to the prolactin regulatory region at several sites and induce transcription [33].

As mentioned above, levels of circulating prolactin increase continuously during the pregnancy period, probably stimulated by rising levels of estrogen, and thought to be involved in the proliferation and differentiation of the mammary cells required for lactation [34]. Estrogens enhance the growth of prolactin-producing cells, stimulate prolactin production directly and also suppress dopamine. Dopamine is the main inhibitory factor of prolactin synthesis as it acts on the dopamine receptor D_2 (D_2R) receptors of lactotrophs and stop prolactin release in the pituitary. During established lactation, milk suckling by the infant stimulates the mechanoreceptors in the nipple, which in turn stimulates the anterior pituitary gland via hypothalamus to synthesize prolactin [2, 6]. Oxytocin, synthesized from the posterior pituitary gland synergistically with prolactin, acts on the breast myoepithelial cells and causes them to contract and eject milk. Synthesis and release of prolactin is in a pulsatile fashion [35, 36]. It has been recently suggested that the presence of a Pit-1 binding site adjacent to E-box133 that binds several circadian elements in the rat prolactin promoter may be necessary for pulsatile gene expression activity [36]. The factor resulting in pulsatile secretion during lactation period is infant suckling.

Prolactin exerts its action via binding to the prolactin receptor (PRLR), which is a transmembrane receptor belonging to the class 1 cytokine receptor superfamily. PRLRs are broadly classified as the long form and the short form receptors [37]. Long PRLR has extracellular domains with two disulphide bridges, which are essential for ligand binding, a transmembrane domain, a duplicated tryptophan-serine motif and also an intracellular signal-transducing domain [32, 38]. Other shorter isoforms are formed as a result of alternative splicing of the long mRNA or through post-translational cleavage of the protein product [37]. The isoforms' extracellular domains

are identical and they vary in the length and composition of their cytoplasmic domains [31]. The canonical isoform is the long one.

Binding of the prolactin to its pre-dimerized receptor induces a conformational change that leads to tyrosine phosphorylation of the Janus kinase 2 (JAK2), which is constitutively associated with the membrane proximal region of the intracellular region of PRLR. Further phosphorylation via JAK2 leads to activation of several signaling cascades [34]. There are three major pathways that prolactin activates through PRLR (Figure 1.3).

First, JAK2 phosphorylates cytoplasmic members of signal transducer and activator of transcription (Stat) family. This initiates the canonical JAK-Stat pathway. The Stat family has eight members. Four of them, Stat1, Stat3, Stat5a and Stat5b have been identified as signal transducing molecules of PRLR [31, 39]. Among these, Stat5a and 5b are recognized as the most important transducers of the long isoform of the PRLR to induce differentiation of mammary cells, milk protein synthesis and proliferation [40]. Stats are present in the cytoplasm as monomers. When activated, they form dimers, translocate into the nucleus and initiate transcription of their target genes by binding to Stat DNA-binding motif in the promoter regions [31]. Stat5a and Stat5b can form homodimers or heterodimers with each other and this represent a major level of differential regulation of transcription initiation by Stat5a and Stat5b [41]. Normal physiological levels of PRL maintain basal activity of Stat5 in virgin mammary glands. Parallel to the changes in PRL synthesis in the body, activation of Stat5 change according during the pregnancy-lactation cycle and involution [42]. One of the well-established target genes of Stat5 is the milk protein gene beta-casein. The beta-casein gene expression has been widely used to monitor the activation of Stat5 upon PRL induction [43].

Secondly, activation of the Src family of tyrosine kinases, especially the proto-oncogene tyrosine-protein kinase Src (c-Src), is required for cell proliferation induced by prolactin. When activated, c-Src is shown to interact with phosphatidylinositide 3-

kinases (PI3K). Activated PI3K then activates protein kinase B (also known as AKT). AKT functions as a serine/threonine –specific protein kinase and regulates cell survival by binding to further downstream elements [44].

Finally activated JAK2 can induce the activation of mitogen-activated protein kinase (MAPK) pathway through SHC/GRB2/SOS/RAS/RAF1 intermediates upstream of MAP kinases. This pathway initiates the transcription of genes associated with proliferation and growth in cells.

In summary, initiation of these three key signalling cascades result in transcription of target genes that cause terminal differentiation, milk protein production, motility, cell proliferation and survival. Alterations in these signalling cascades can lead to abnormal proliferation of epithelial cells and contribute to breast tumorigenesis. Supporting a positive link between prolactin signalling and breast cancer, epidemiologic studies indicate that postmenopausal women with 'high-normal' levels of prolactin are at increased risk of breast cancer compared to women with 'normal' levels of prolactin [45, 46]. Also, mutations such as ligand-independent-dimerization of PRLR are detected in breast tumours [47]. Likely, in a study involving 70 multiple fibro adenoma patients (benign breast tumours), four patients were identified having a heterozygous substitution mutation in the extracellular domain of PRLR gene. This results in; prolactin-independent tyrosine phosphorylation, activation of Stat5 signalling and increased cell proliferation and protection from apoptosis [48]. Overall, prolactin-PRLR signalling promotes normal and malignant breast growth through a number of pathways.

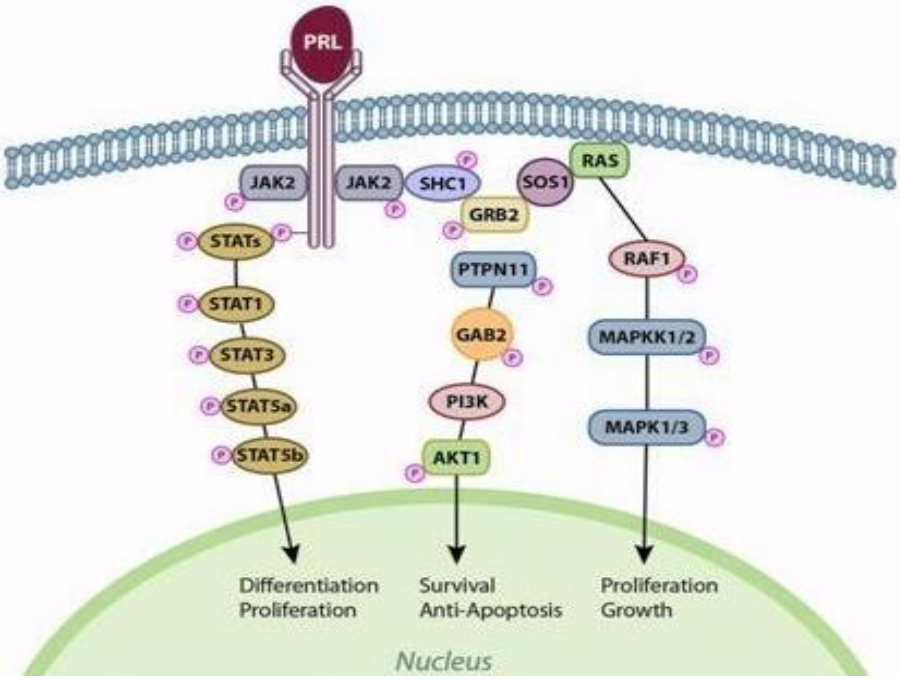

Figure 1.3. Prolactin induced signaling pathways.
When bound to its receptor, prolactin causes phosphorylation of Janus kinase 2 (JAK2) protein. Phosphorylated JAK2 then induces activation of Stat5a/b, phosphatidylinositide 3-kinases (PI3K) and mitogen-activated protein kinase (MAPK) via phosphorylation. Downstream signaling pathways are then induced, which results in transcriptional activation of genes involved in differentiation, proliferation, survival, anti-apoptosis and cell proliferation. Figure is taken from [49].

As mentioned earlier, understanding the regulation and effects of prolactin *in vivo* is complex. One of the complexities is caused by the ability of human PRLR to bind at least three ligands including prolactin, placental lactogen and growth hormone [50]. Also, different isoforms of PRLR can possess different signaling properties. For example, short PRLR is not tyrosine phosphorylated and this prevents it from interacting directly with the Stat factors [51]. On the other hand, the 16kDa prolactin isoform lacks the C-terminal region that is required for PRLR binding and it can induce different signaling pathways via binding to different receptors. As a result it has been shown that it can exert inhibitory effects on angiogenesis and tumorigenesis, unlike the major 23kDa isoform [52].

Second signalling pathway that is well established in breast physiology is the ERBB induced signalling. ERBB family of receptors include four closely related receptor

tyrosine kinases, which are EGFR (HER1, ERBB1), HER2/c-neu (ERBB2), HER3 (ERBB3) and HER4 (ERBB-4). Epidermal growth factor receptor (EGFR, ERBB1, HER1) is essential for mammary ductal growth and branching morphogenesis. EGFRs can be activated through binding of its ligands; epidermal growth factor (EGF), transforming growth factor α (TGFα) and amphiregulin (AREG), that all have roles in epithelial cell proliferation or differentiation processes. Cripto1 (CR1), for example, is a novel EGF-related protein that triggers the branching in mammary epithelium and inhibits the expression of milk proteins during pregnancy. CR1 was also found to be overexpressed in breast tumours [26, 53]. EGFR is upregulated at nearly half of triple negative breast cancers [54]. Recent studies demonstrated that EGFR and its downstream pathways can regulate epithelial-mesenchymal transition in breast tumours. Also, the EGFR pathway is implicated in angiogenesis and cell invasion by its regulation of the expression and activity of matrix metalloproteinases (MMPs) [55, 56].

ERBB2 (HER2) is amplified in 25 to 30% of breast cancers and in these cases the encoded protein is present in abnormally high levels in the malignant cells. Amplification of the receptor increases the aggressiveness of the tumour and decreased overall survival of the patients [57]. HER2 protein activates the Ras-MAPK pathway and inhibits cell death through the PI3K/Akt-mammalian target of rapamycin (mTOR) pathway [58]. Therefore, strategies to target HER2 have been of a great importance in treating breast cancers. One such medication is trastuzumab (Herceptin), which is a humanized monoclonal antibody that binds to the extracellular domain of HER2. When bound, it inhibits the proliferation and survival of HER2-dependent tumours [58]. The overexpression of the EGFR and the ERBB2 receptors are also associated with progression to hormone independence in human breast cancer [59].

ERBB4 (HER4) is also an essential factor for mammary gland differentiation. It mediates the Stat5 signalling by directly activating Stat5A through phosphorylation of the regulatory Tyr-694 or through serine phosphorylation at mid-lactation [60]. This

activation of Stat5 is independent of prolactin-induced activation and how these two pathways intersect, if they do, is unclear [61]. It was shown that ERBB4 signalling enhances growth of many human breast cancer cells and promotes murine epithelial cells to form tumours [25].

Third important signalling pathway in breast development and tumorigenesis is activated by insulin-like growth factor 1 (IGF1). It is the requisite factor for mammary gland development causing proliferative and anti-apoptotic events [5]. IGF1 expression is induced when growth hormone (GH) binds to the GH receptors in the stroma. Then, IGF1 acts on its receptor (IGF1R) in the epithelium and induce ductal development [5, 62]. In a study it was shown that mice, which lacks IGF1 or IGF1R, the mammary glands lack terminal end buds and exhibit diminished ductal outgrowth [63, 64]. Aberrant expression of each component of IGF1 signalling system, which include IGF1, IGF1R and IGF binding proteins (IGFBPs), was reported in breast cancers [65]. It has been shown in a study that higher circulating IGF1 levels can increase the risk of breast cancers diagnosed at 50 years or older [66]. In another study, IGF1 was shown to stimulate proliferation and bone localization of breast cancer cells *in vivo* through activation of AKT and recruitment of transcription factor NFkB [65, 67]. As expected, IGF1 receptors are found to be frequently overexpressed in breast cancer patients [65]. Overexpression of IGF1R is implicated in breast cancer by increasing the metastatic potential of the original tumour by inferring the ability to promote vascularisation. It was also shown that when EGFR inhibitors are being used to inhibit the EGFR signalling pathway, IGF1R could confer resistance by forming one half of a EGFR heterodimer [56]. However, in contrast, no correlation was detected between IGF1 levels and breast cancer development in women during early pregnancy [68]. IGFBPs, on the other hand, can have either stimulatory or inhibitory effects on IGF-induced pathways. IGFBP5, for example, inhibits migration of MCF7 cells by inducing cellular adhesion via activation of AKT in an IGF-independent manner [65, 69]. In contrast, some findings suggest a significant association between systemic and tissue IGFBP3 levels and poor breast cancer prognosis especially in estrogen receptor (ER) positive breast tumours [70, 71].

Finally, fourth key signalling cascade involved in breast tumorigenesis as well as normal breast development is the estrogen (E2), estrogen receptor (ER) system. Role of E2 in mediating breast development during puberty and pregnancy in preparation of lactation is crucial. It is especially involved in ductal development, causes adipose tissue deposition and also takes part in connective tissue growth. Estrogen exerts its actions in conjunction with the growth hormone and its secretory product IGF-1 [72, 73]. When bound to its receptor, they form a complex and translocate into the nucleus, where ER acts as a transcription factor via binding to estrogen response elements (EREs) and initiate transcription of genes involved in cell growth and proliferation [74, 75]. Also, it increases the progesterone receptor expression in breast and induces the secretion of prolactin hormone [76, 77]. Progesterone receptor upregulation is also used as a prognostic marker for breast tumorigenesis [78].

Therefore, exposure to estrogen may be implicated in increasing breast cancer risk because of its role in inducing cell proliferation and its effect on expression of other hormones or receptors that stimulate cell division. Its receptor, ER, is known to be upregulated in malignant breast tissue and is long been used as a major prognostic marker [79]. Breast tumours are classified as hormone receptor positive tumours in case they overexpress the ER. Over activity of the receptor causes aggressive proliferation of breast tissue, enhancing tumour growth, hence inhibition of the ER has become one of the major strategies for the prevention and treatment of breast cancer [78]. ER is one of the few tumour markers, which is approved for routine clinical use by the American Society of Clinical Oncology [80]. The first targeted therapy that was invented against breast cancer, tamoxifen (ICI 46,474) targets the ER [81]. Tamoxifen competitively binds to the ER and inhibits estrogen action. Therefore, expression of estrogen-regulated genes including angiogenic factors and growth factors are inhibited leading to the antitumor effects of tamoxifen [82].

In summary, regulation of growth factors and receptors as well as their downstream signalling elements is crucial in regulation of normal breast development and neoplastic transformation of the tissue. It is also of a great importance to understand

that with the advancements in genomic technologies, we have a better understanding of molecular subtypes of breast cancer enabling us to target expression of specific receptors (ER and HER2) in certain subtypes by drugs that improve the mortality and effective morbidity rates [83]. However it is also clear that some aspects of breast cancer tumorigenesis are not completely known, such as the lack of targeted therapies for TNBCs and development of drug resistance. Therefore investigating mutations and mechanisms in TNBCs gains more importance in terms of drug development.

To begin addressing these issues, understanding normal breast mammary development and differentiation is impeccable. To contribute to these efforts, we consider alternative polyadenylation (APA) to be an important aspect of post-transcriptional gene regulation that may explain some mechanistic aspects of both mammary differentiation and breast tumorigenesis.

1.3. Alternative Polyadenylation (APA)

3’ ends of all eukaryotic nascent mRNAs are processed to become mature mRNAs. Processing involves splicing, capping, editing and polyadenylation of the premature mRNA molecules. Defects in pre-mRNA 3’ processing have been associated with a wide spectrum of human diseases so it is important to understand the mechanisms and regulation involved [84]. We are particularly interested in understanding regulation of polyadenylation event and the concept of alternative polyadenylation in breast development and cancer.

Coupled to transcription, the newly synthesized pre-mRNA undergoes endonucleolytic cleavage at a polyadenylation site determined by binding of polyadenylation factors and a string of adenosine bases are added to the upstream cleavage fragment by poly (A) polymerase (PAP) enzyme [85]. Addition of the poly (A) tail protects the mRNA from exosome nucleases by protecting the 3’ downstream sequences and prolongs the lifespan of mRNA, increasing its stability [86]. It can also influence transport, localization and translation of the mature mRNA [87].

The canonical poly (A) signals are defined by multiple cis-elements involving an AAUAAA hexamer located 10-30 nucleotides upstream of the poly (A) site, a GU- or U- rich downstream element (DSE) located ~30 nucleotides of the poly (A) site and other U-rich upstream stimulatory elements (USEs) [84, 85] (Figure 1.4). When the trans-acting elements bind to their target cis-elements, pre-mRNA 3' processing complex assembly is initiated. The complex consists of the poly (A) polymerase (PAP), the poly (A)-binding proteins (PABPs), the RNA polymerase II large subunit (RNAP II) and four multi-subunit protein complexes; CSTF (cleavage stimulation factor), CPSF (cleavage and polyadenylation specificity factor), CFIm (cleavage factor Im) and CFIIm (cleavage factor IIm) (Figure 1.5) [84]. Most of these factors and their subunits are highly conserved through mammals, yeast and plants.

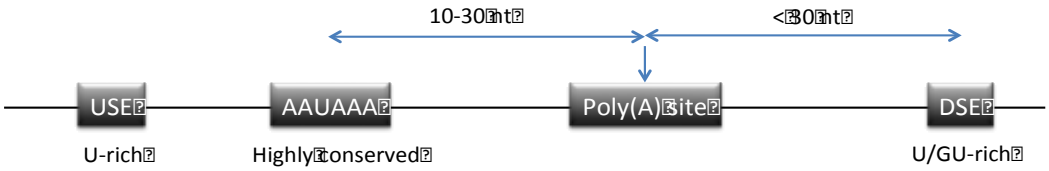

Figure 1.4. Schematic representation of polyadenylation elements on an mRNA.
There are GU or U-rich downstream elements (DSE) located more than 30 nucleotides downstream of the poly(A) site. Upstream elements include the highly conserved AAUAAA hexamer about 30 nucleotides upstream of the poly(A) site and U-rich upstream stimulatory elements (USEs).

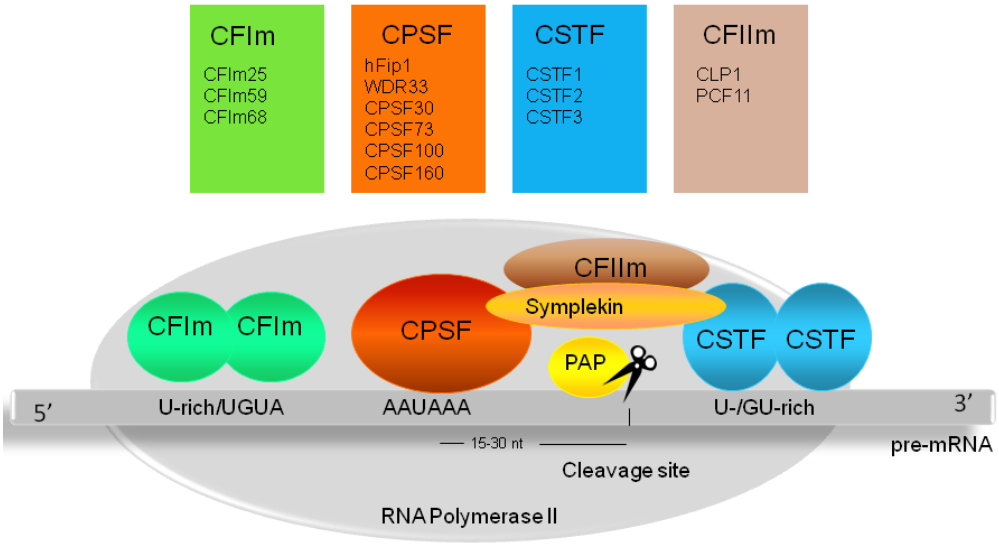

Figure 1.5. Schematic representation of the main factors involved in cleavage and polyadenylation complex.
There are several proteins involved in the cleavage and polyadenylation processes in the cell. CPSF, CSTF, CFIm and CFIIm are required for cleavage. Poly(a) polymerase (PAP) and poly(A)-binding proteins (PABs) are crucial for pre-mRNA cleavage and polyadenylation. Cleavage sites are represented by the poly(A) signals (PAS). Protein complexes bind to the upstream and downstream elements (USE and DSE) to regulate the cleavage and polyadenylation processes. Figure is taken from [88].

CPSF is the key protein complex that is required for cleavage and polyadenylation of the precursor mRNAs. It consists of CPSF160, 100 and 73 subunits. It recognizes the A(A/U)UAAA signal, helps to recruit other factors of the 3' processing complex and catalyses cleavage [84]. CSTF complex, on the other hand, recognizes the DSE and it is a hetero-trimeric protein composed of the CSTF77, CSTF50 and CSTF64 subunits [89]. Stable association of CSTF to DSE is known to require cooperative binding of CPSF to the A(A/U)UAAA element [84]. Binding of the CSTF complex to RNA is mediated by the CSTF-64 subunit and this interaction is very strong [89]. CFIm is involved in cleavage step of pre-mRNA processing. It consists of three subunits as CFIm25, CFIm59 and CFIm68. It is reported that all three subunits could be cross-linked to RNA, suggesting that they are all involved in RNA binding [84, 89]. Even though CFIm was able to stabilize CPSF complex on RNA, CPSF did not enhance the binding of any CFIm subunit, nor did mutation of AAUAAA affect their binding efficiency [89]. Further investigation revealed that CFIm25-68 binds specifically to

UGUAN motif, mediated by the CFIm25 subunit [84]. CFIIm, like CFIm, is involved in the cleavage of pre-mRNA, however its exact function is not yet clear. Two subunits of this complex have been identified in human cells. These are hClp1 and hPcf11. It is thought that CFIIm forms a bridge between CFIm and CPSF complexes as it interacts with both [90].

It has been discovered that a great proportion of mammalian genes have more than one functional poly (A) site. More specifically, in a study the percentage of genes that undergo alternative polyadenylation (APA) was shown to be more than 70% in humans and ~32% in mouse [87, 91]. Selection of different poly (A) sites can result in the production of different transcript variants from a single gene. Generally alternative polyadenylation events can be divided into three groups [92]. In the first group, there is only one poly (A) signal in the 3' UTR, therefore only one mRNA isoform is produced with only one possible protein product. In the second group, polyadenylation event is coupled to splicing. In this case alternative poly (A) sites can be inside the exons or introns. Therefore the coding sequence may be altered according to the position of the stop codon and different/altered protein products can be produced [93]. In the last group, there are more than one poly (A) sites in the 3' UTR region of the gene. Selection of one of these poly (A) sites results in multiple mRNA isoforms differing in their 3' UTR length. Even though the poly (A) tail locations of these transcripts are different, the coding sequence is not altered so the same protein is translated [93]. However, presence or absence of regulatory regions on the 3'UTR can eventually affect the protein levels. More cis-elements such as microRNA-binding sites and RNA binding protein regions are eventually present in long 3' UTR regions. Binding of such elements to the mRNA can affect its localization, stability and translation [93, 94]. Therefore, translation efficiency of a transcript can partially be determined by the selected poly (A) tail location.

1.4. APA In Cancer and Development

Recently, it has become increasingly evident that APA plays key role during gene expression regulation, transformation, development, differentiation and cell proliferation [93]. One of the evidence on global APA dependent gene expression was shown in primary murine T cells that express CD4+ and cells stimulated through TCR (T cell antigen receptor). Increased expression of shorter 3'UTR isoforms were observed in stimulated cells [94]. Further investigation of APA events in cell lines revealed a link between increased proliferation rates with shorter 3'UTR isoforms. At the protein level, it was shown that shorter 3'UTR isoforms yielded more protein products when compared to full lengths [93-95]. This supports the notion that binding of more microRNAs on longer isoforms decreases the stability of mRNAs resulting in decreased protein amount in rapidly dividing cells. Data obtained from RNA sequencing work has been widely used to detect APA events during proliferation and cancer. Increased protein levels of oncogenes that have shortened 3'UTR transcripts suggest a pivotal role of APA in cancer [95].

In development and differentiation, on the other hand, 3'UTR lengthening of genes transcripts seems to have a vital role [93]. For example, progressive lengthening of 3'UTRs in mouse embryonic development was detected using EST (expressed sequence tag) and SAGE (series analysis of gene expression) data. It is thought that lengthening of the untranslated regions reveal more cis elements, such as AREs, which can affect gene regulation and coordinate their activities with development [96].

1.5. Aim of The Study

As APA is emerging as a novel post-transcriptional regulator of gene expression, we hypothesized that APA may also be regulating gene expression during mammary developmental stages. Our aim was to identify APA regulated transcripts in the P-L cycle of mammary tissue that may have further roles in breast cancers.

CHAPTER 2

MATERIALS AND METHODS

2.1. Microarray Dataset

Microarray dataset was obtained from Prof. Christine Watson's website [97]. The data contains expression levels for 12,088 mouse transcripts across 12 time points of the mouse mammary pregnancy cycle [98]. Specimens were taken from the abdominal mammary glands of virgin, pregnant (gestation day 5, 10, 15), lactating (day 0, 5 and 10) and involuting (hours 12, 24, 48, 72 and 96) animals. There are 2 biological replicates from each timepoint [98].

2.2. APADetect

APADetect is a probe level screen algorithm, which was developed by Dr. Tolga Can (Depertment of Computer Engineering, METU, Ankara). The tool is used to detect expression levels of different 3'UTR isoforms using the available microarray data [93, 99]. In the MG-U74A2 array, each of the 12,088 murine transcripts are represented by 16 probe pairs. The 16 probes/transcript can be divided into proximal and distal probe sets based on the poly (A) site locations taken from the PolyA_DB. The abundance of "short" and "long" 3'UTR isoforms are determined by analysing mean signal intensities of proximal and distal probe sets for each gene. For each transcript, mean of proximal probes intensity is divided by the mean of distal probes intensity to calculate the short to long ratio (SLR) value and the values of different timepoints are compared to each other to detect any APA based changes. SLR values from lactation, gestation and involution timepoints were normalized to that of virgin timepoint.

2.3. Significance Analysis of Microarrays (SAM)

To further analyse the APADetect output file, TIGR Multiexperiment Viewer (MeV) software was used (http://www.tm4.org/) [100]. MeV is a microarray data analysis program for visualization, clustering, statistical analysis, classification and biological theme discovery.

Significance analysis for microarrays (SAM) in the MeV software is a statistical method that was designed to detect significant genes in a set of microarray data, taking into account t-tests specific for the genes. A score is assigned to each gene according to the change in gene expression levels, relative to the standard deviation of different measurements. The scores above a threshold level are assumed as statistically significant [100].

In our case, log SLR data were determined for each gene in the array as the output of APADetect analysis. Then SAM was used to analyse individual SLR values for statistical significance. Repeated permutations of the data are used to determine if the SLR values of any gene are significantly different than the control samples. The statistically significant genes were listed according to positive and negative significance, which indicates shortened or lengthened genes respectively. Then these genes were individually analysed.

2.4. Ontology Analysis

For the ontology analysis DAVID (The Database for Annotation, Visualisation and Integrated Discovery) was used. DAVID is a tool designed to classify genes according to their biological process, molecular function, pathway or family/subfamily [101, 102]. In this work, genes that have log SLR values greater than 1.5 and less than 0.6 were clustered according to their GOTERM biological process. Clusters that have enrichment scores above 1.25 and p-values below 0.05 were considered as significantly enriched within the given set of genes [102].

2.5. Cell Lines, Cell Culture Conditions and Treatments

4T1 cell line was a kind gift from Dr. Özgür Şahin, Bilkent University, Ankara, Turkey. 4T1 cells were grown in DMEM with Earle's salts (Biochrom) supplemented with 10% fetal bovine serum (FBS), 1% Penicillin/Streptomycin (P/S) and 1% non-essential amino acids (NEAA).

HC11 cell line was a kind gift from Prof. Jonas Tallkvist, Swedish University of Agricultural Sciences, Uppsala, Sweden and was cultured by permission of Prof. Bernd Groner, Institute for Biomedical Research, Frankfurt, Germany and Dr. *Nancy Hynes*, Friedrich Miescher Institute, Basel, Switzerland. HC11 cells were grown in RPMI 1640 medium containing L-glutamine and 20 mM HEPES (Biochrom) supplemented with 5 μg/ml bovine insulin, 50 μg/ml gentamicin, 10 ng/ml epidermal growth factor (EGF), 7.5% $NaHCO_3$, and 10% FBS. Cell lines were expanded and incubated as monolayers in T75 tissue culture flasks at 37^0C with 5% CO_2 and 95% humidified air.

For prolactin-induction experiments, HC11 cells were cultured to confluency. 48 hours post-confluency, cells were cultured for an additional 24 hours in the FBS-containing RPMI 1640 medium without EGF. Differentiation of the HC11 cells was then achieved by culturing the cells for 6, 12, 24, 36, 48 and 72 hours in EGF- and FBS-free RPMI 1640 medium supplemented with 1 μg/ml, 2 μg/ml, 3 μg/ml, 4 μg/ml and 5 μg/ml prolactin (Sigma-Aldrich, Cat no. L6520) and 1 μM hydrocortisol (Sigma-Aldrich, Cat no. H0888). The treatment conditions were then optimized as 36 hours with 5μg/ml prolactin concentration. Undifferentiated HC11 cells were cultured in parallel with the FBS- and EGF-containing RPMI 1640 medium as untreated controls.

Cell freezing of both 4T1 and HC11 cells in liquid nitrogen was done when the cells reached 80% confluency. Cells were pelleted by centrifuging at 1400 rpm for 5 minutes and resuspended in 1 ml media containing 10% dimethyl sulfoxide (DMSO) for long-term storage in liquid nitrogen.

2.6. RNA Isolation

Total RNA isolation was done using 5 PRIME PerfectPure™ RNA Purification System. Experiments were performed according to the manufacturer's instructions.

2.7. DNase I Treatment

All the solutions in DNase treatment procedure was prepared using DEPC-treated water. To obtain DNA-free RNA, isolated RNA samples were treated with Deoxyribonuclease I (DNase I) from Fermentas (Cat no EN0521). Components of the reaction mixture are listed in Table 2.1. below.

Table 2.1. DNase treatment reaction mixture.

RNA	15 µg
10X Reaction Buffer	10 µl
DNase I (1u/µl)	3 µl
DEPC water	Variable
TOTAL	100 µl

The mixture was prepared on ice and the tubes containing the mixture were incubated at 37^0C waterbath for 60 minutes. 100 µl Phenol: Chloroform: Isoamyl Alcohol (25: 24: 1) was used to stop the enzymatic reaction. Mixture was mixed by vortexing for 30 seconds, all mixture was transferred to phase lock tubes and incubated on ice for 10 minutes. Then the samples were centrifuged for 20 minutes at 14000 g at 4^0C. After centrifugation, the upper phase that contained RNA was taken to a fresh tube. This phase is approximately 80 µl. Then 240 µl 100% ethanol and 8 µl 3M NaAc was added onto the RNA containing samples and were left for incubation at -20^0C overnight. After incubation, samples were centrifuged for 30 minutes at 14000g at 4^0C. Supernatant was discarded and the RNA pellet was washed with by 600 µl 70% ethanol and centrifuged again for 15 minutes at 14000g at 4^0C. The supernatant was discarded and the dry RNA pellet was dissolved in 25µl MG Water.

Absence of DNA contamination was confirmed by conventional PCR using

cyclophilinB (CycB) specific primers. In the presence of DNA contamination, the primers produce a 306 base pair product. PCR reactions were done using following conditions: incubation at 94^0C for 10 minutes, 35 cycles of 94^0C for 30 seconds, 60^0C for 30 seconds and 72^0C for 30 seconds and final extension at 42^0C for 5 minutes.

2.8. RNA Quantity and Quality Determination

RNAs were quantified via BioDrop µLITE. A260/A280 and A260/A230 ratios were used to determine the purity of RNA samples. For all RNA samples, A260/A280 ratio was between 1.8-2.00 and A260/A230 ratio was 1.8 and higher.

2.9. cDNA Synthesis

cDNA synthesis was done by using the RevertAid First Strand cDNA Synthesis Kit (Cat no 1622; Life Technologies). Reaction conditions are described and listed in Table 2.2.

Table 2.2. Reverse transcription reaction conditions.

RNA	1000 ng
Primer (oligodT)	1 µl
MG Water	Variable
TOTAL	12 µl
Centrifuged and incubated at 70^0C for 5 mins.	
5X Reaction Buffer	4 µl
Riblock RNAse inhibitor	1 µl
dNTP mix	2 µl
RevertAid RT enzyme	1 µl
TOTAL	20 µl
Briefly centrifuged and incubated at 42^0C for 60 minutes; reaction was then stopped by heating up to 70^0C for 5 minutes.	

2.10. β-Casein Expression Analysis

When HC11 cells are treated with prolactin hormone, the expression of the major milk protein β-casein (*Csn2*) (accession: NM_009972.2) is known to be induced [103]. Expression of β-casein was quantified by RT-qPCR as a positive control for the prolactin treatment. Primers used in PCR reactions were β-casein_F: 5'-CTTAACCCCACCGTCCAAT-3' and β-casein_R: 5'-AGCATGATCCAAAGGTGAAAA-3'. Following conditions were used in RT-qPCR reactions: incubation at 94^0C for 10 minutes, 35 cycles of 94^0C for 30 seconds, 62^0C for 30 seconds and 72^0C for 30 seconds (plate read). Incubation at 95^0 C for 10 seconds and melt curve from 50^0C to 99^0C with 1^0C, 0:05 increment (plate read).

2.11. Quantitative RT-PCR Analysis

For quantitative RT-qPCR, SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad Laboratories) was used using the CFX Connect™ Real-Time System (Bio-Rad Laboratories). 3µl of synthesized cDNA was used in 10 µl reaction samples containing 5X Universal SYBR® Green Supermix buffer and 300 nM of gene specific primer pairs. MIQE guidelines were followed throughout the PCR and RT-qPCR analyses [104]. Fold change of the gene expressions were normalized against the reference gene, cyclophilinB (*Ppib*, *CycB*) (accession: NM_011149) as its expression is consistent in all mouse cell lines used and does not change in response to prolactin treatment [105]. Reaction efficiency correction and ΔΔCq method was used for quantification [106]. Sequences of all mouse primer pairs, expected product sizes in base pairs and RT-qPCR annealing temperatures are summarized in Table 2.3.

Table 2.3. Primer sequences, product size and annealing temperatures for mouse cDNA.

Gene Accesion Number	Primer name	Sequence (5'-3')	Product size	Annealing Temp.
Thrsp (NM_009381.3)	Thrsp short F	ACGGAGCCCCTGATCTCTAT	194bp	57^0C
	Thrsp short R	GGCTTCTAGGTCCAGCTCCT		
	Thrsp long F	CGCAGCTCTTGCTTTTATCC	108bp	57^0C
	Thrsp long R	CAATAGTTAGGTTCATGTGCT		
Sdc2 (NM_008304.2)	Sdc2 short F	CAAGACCAAAGAACCCCTCA	226bp	60^0C
	Sdc2 short R	TCGCTTTGCTTCTTTGGAAT		
	Sdc2 long F	GCACAATGCCCCTCAATTAC	156bp	60^0C
	Sdc2 long R	TGTGCTTTACAGCAGTGACAA		
Xdh (NM_011723.3)	Xdh short F	TCCAGCTAACGTCCAGCTTT	218bp	62^0C
	Xdh short R	TGGCTTCTGAAGTGTCGATG		
	Xdh long F	TTCTCACACACCTCCTGACG	153bp	62^0C
	Xdh long R	CCCACACACACACACACACT		
Igfbp7 (NM_001159518.1)	Igfbp7 short F	GGAGGACGCTGGAGAGTATG	177bp	67^0C
	Igfbp7 short R	GGCTGTCTGAGAGCACCTTT		
	Igfbp7 long F	ATTCGCACCAAGGAAATCTG	156bp	65^0C
	Igfbp7 long R	TGGTCTGAGAGCCTGCACTA		
Hnrnpa1 (NM_010447.5)	Hnrnpa1 short F	CACGGAACCAAGGTGGCTAT	168bp	66^0C
	Hnrnpa1 short R	TGTGCTTGGCTGAGTTCACA		
	Hnrnpa1 long F	CAACTGTATGGATTTGGGACT	150bp	60^0C
	Hnrnpa1 long R	TCTCTGGACTTATAGGTGGCA		
Cstf2 (NM_133196.6)	Cstf2 F	CCTACCAACGTCCCAACTCC	206bp	56^0C
	Cstf2 R	AGGGGTCCTCCTCTCATGTC		
CFIm25 (NM_026623.3)	CFIm25 F	CTCGACTGTTAATGGCGGGT	208bp	60^0C
	CFIm25 R	GGCTTGGTCTGCTGGATGTA		
Csn2 (NM_009972.2)	β-casein F	CTTAACCCCACCGTCCAAT	143bp	62^0C
	β-casein R	AGCATGATCCAAAGGTGAAA		
CycB (NM_011149)	CyclophilinB F	GCGCAATATGAAGGTGCTCT	304bp	60^0C
	CyclophilinB R	GAAGTCTCCACCCTGGATCA		

2.12. Total Protein Isolation

Total proteins were isolated using the M-PER Mammalian Protein Extraction Reagent (Cat no 78501; Thermo Scientific). Cells were washed with PBS twice. M-PER was added directly on monolayer cells (250 µl for 60 mm culture dishes), shaken on shaker for 15 minutes on ice. Cells were then transferred into an eppendorf tube using a scraper. Then samples were centrifuged for 15 minutes at 14000g at 4^0C. Supernatants, which contain the total protein, were then transferred into a fresh eppendorf tube and stored at -80^0C.

2.13. Western Blotting

Denaturation of the total protein extract (70 µg) was done by 6X Laemmli buffer at 100^0C for 10 minutes. Then samples were run on a 7.5% polyacrylamide gel. Then they were transferred onto a nitrocellulose membrane. For detection of the Cstf2 protein, membrane was blocked with 5% non-fat milk in TBS-T (Tris Buffered Saline-Tween) and left for overnight incubation shaking at 4^0C with the monoclonal anti-Cstf2 rabbit antibody (1:100 dilution; ab64942, Abcam). Then the membrane was incubated with the secondary anti-rabbit antibody (1:2000; Santa-Cruz Biotechnology) for 1 hour at room temperature and visualized.

For CFIm25 protein, membrane was blocked with 5% non-fat milk in TBS-T (Tris Buffered Saline-Tween) and left for overnight incubation shaking at 4^0C with the monoclonal anti-CFIm25 rabbit antibody (1:100 dilution; ab64942, Abcam). Then the membrane was incubated with the secondary anti-rabbit antibody (1:2000; Santa-Cruz Biotechnology) for 1 hour at room temperature and visualized.

To detect Hnrnpa1 protein, membrane was blocked with 5% BSA in TBS-T (Tris Buffered Saline-Tween) and left for overnight incubation shaking at 4^0C with the monoclonal anti-Hnrnpa1 rabbit antibody (1:750 dilution; ab177152, Abcam). Then the membrane was incubated with the secondary anti-rabbit antibody (1:2000; Santa-Cruz Biotechnology) for 1 hour at room temperature and visualized.

For protein loading control, anti-β-actin mouse antibody (1:1000; sc-47778, Santa-Cruz Biotechnology) was used. Blocking was done with 5% bovine serum albumin (BSA) in TBS-T. After incubation with the primary antibody, membrane was then incubated in 1:2000 diluted anti-mouse antibody and visualized.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. APA Events in Mammary Gland Development

To investigate whether APA patterns are developmentally regulated in the mammary tissue, we re-analysed a gene expression dataset collected from virgin, pregnant, lactation and involution stages of mouse mammary tissue. The dataset contains expression values of 12,088 mouse mRNAs obtained from specimens taken from abdominal mammary glands of virgin, pregnant (gestation day 5, 10, 15), lactating (day 0, 5 and 10) and involuting (hours 12, 24, 48, 72 and 96) mice.

APADetect analysis of all timepoints compared to the virgin animal data revealed around 2600 APA cases with shortening, lengthening and constant SLR values in total. APADetect results of SLR values were then analyzed using SAM to detect statistically significant events. A total of 80 genes were detected to be significantly shortened or lengthened. Out of these 80 APA events, 50 of them were identified as 3'UTR shortening and other 30 of them were 3'UTR lengthening events (Figure 3.1).

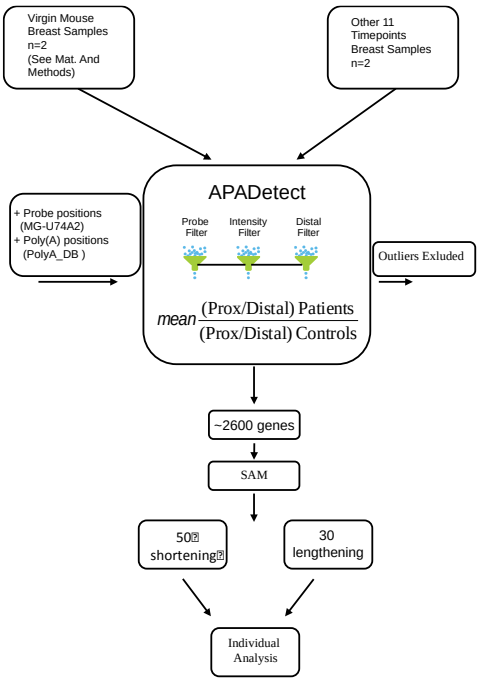

Figure 3.1. Workflow chart.
Steps of APADetect analysis to determine shortened and lengthened mRNAs' SLR values.

Next, we investigated the nature of 3'UTR shortening and lengthening events in gestation, lactation and involution steps (Figure 3.2). There is a clear accumulation of 3'UTR shortening events at gestation and lactation timepoints. However, almost all lengthening events are seen in late lactation and involution timepoints.

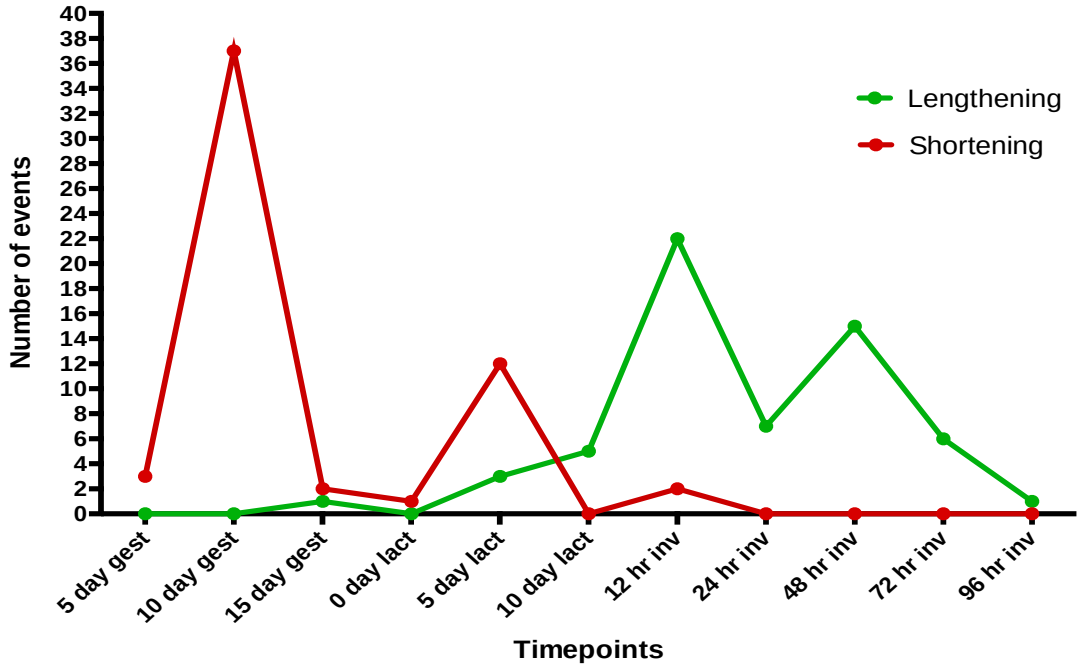

Figure 3.2. Significantly shortened and lengthened gene numbers at each timepoint. Number of genes that are significantly shortened (red) and lengthened (green) at each timepoint, compared to virgin were represented in this graph. Almost all shortening events accumulated at gestation and lactation timepoints, whereas lengthening events were almost exclusively observed during involution.

In addition, we observed a dynamic pattern of APA across different timepoints (Figure 3.3). According to the changes in their log SLR values, we clustered APA events into four groups. Two groups within the shortened genes (a,b) and two groups within the lengthened genes (c,d). Genes in group (a) have high SLR values in virgin, becomes higher during gestation lactation and then somewhat returns to lower levels in involution. However, genes in group (b) have SLR values below 1.0 in virgin then increase significantly at early gestation timepoints. Recovery of SLR values of genes in this group at the end of involution is more obvious.

Other two groups are within the lengthened genes. In group (c) the SLR values of genes are high in virgin and gestation timepoints and we can observe the lengthening events during late lactation and involution. Genes in group (d), on the other hand, have higher SLR values in virgin, gestation and lactation and they become lengthened

during early involution timepoints and recover at the end of involution. Log SLR graphs of 6 representative genes, two from each cluster, are shown in (Figure 3.4).

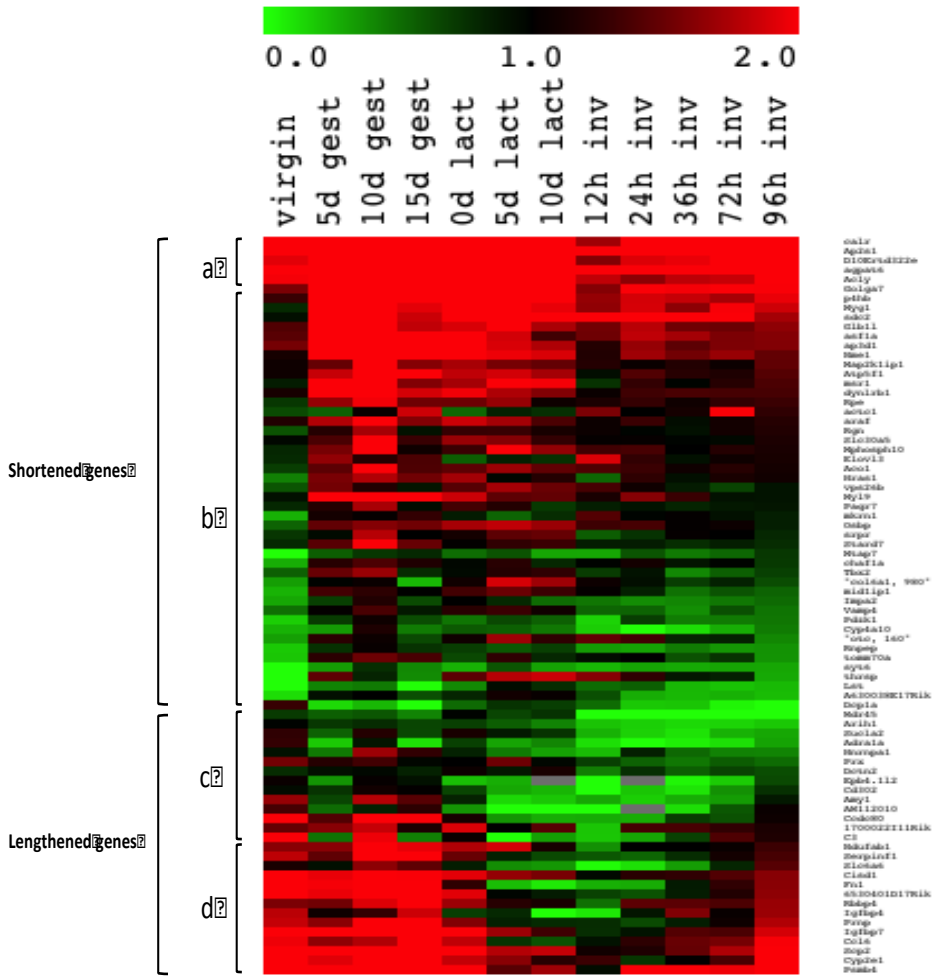

Figure 3.3. Heatmap showing log SLR values of significantly shortened and lengthened genes at all timepoints.
Log SLR values of 50 significantly shortened and 29 significantly lengthened genes, compared to virgin, at all timepoints are represented in the heatmap generated by using the MeV software [100]. The color scale above represents colors of the corresponding log SLR values. Genes were clustered according to the changes in their SLR values. Cluster (a) represents genes that have high SLR values through all timepoints. Genes in cluster (b) have SLR values between 0.0-1.0 in virgin timepoint and undergoes significant shortening at the beginning of gestation, then their SLR values drop back to the levels similar to that of virgin timepoint. Genes in cluster (c), have SLR values between 1.0-2.0 in virgin and gestation timepoints, then undergo significant lengthening during involution. Finally cluster (d) represents genes that undergo significant lengthening during late lactation and early involution and recover at the end of involution timepoints.

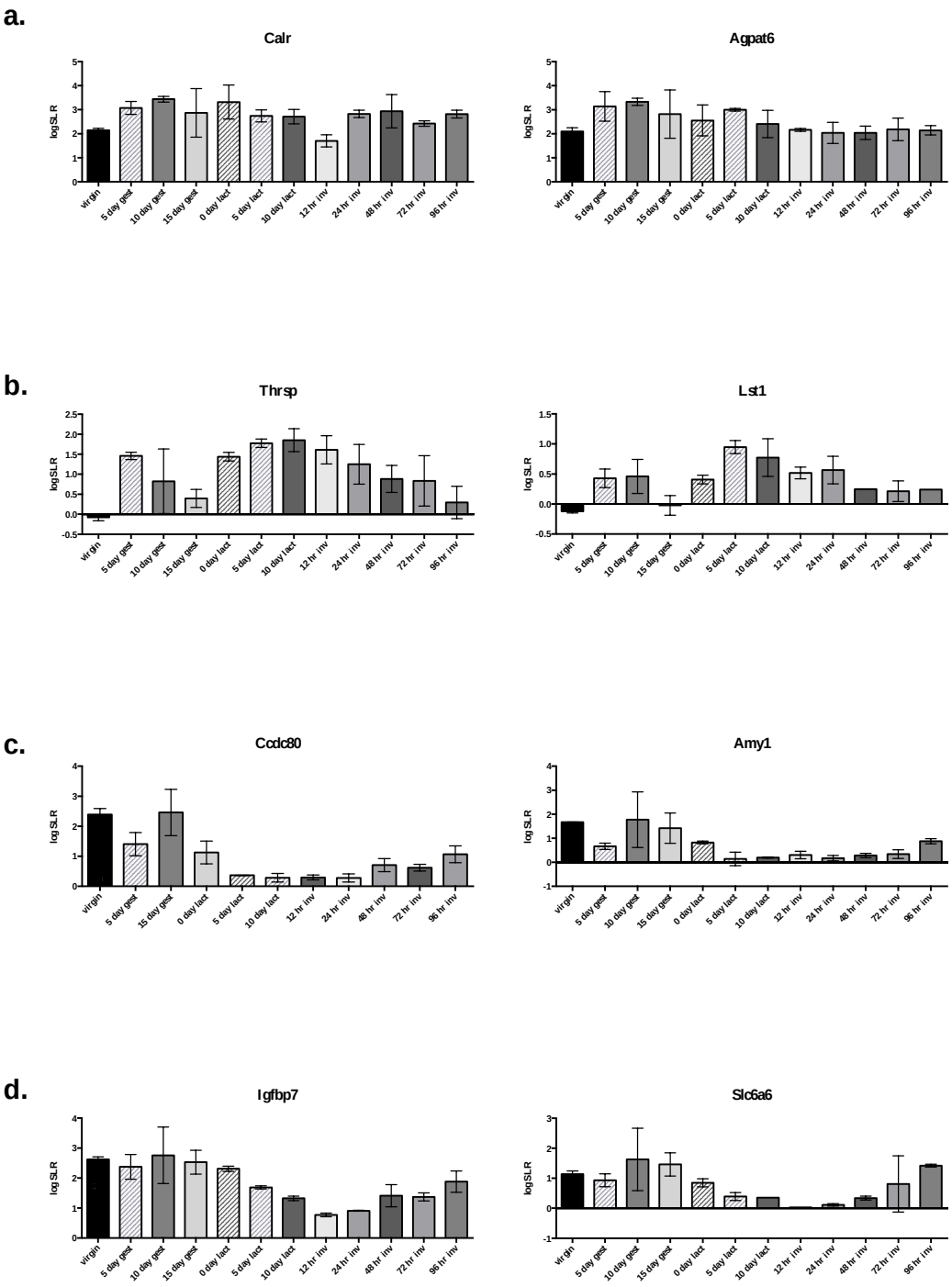

Figure 3.4. Log SLR graphs of eight genes, two from each cluster shown in the heatmap. Genes were clustered according to changes in their log SLR values in (Figure 3). Two genes from each cluster were selected to represent log SLR graphs separately. Calreticulin (Calr) and 1-acylglycerol-3-phosphate O-acyltransferase (Agpat6) are genes from cluster (a). Thyroid hormone responsive gene (Thrsp) and leukocyte specific transcript 1 (Lst1) genes from cluster (b). Coiled-coin domain containing 80 (Ccdc80) and amylase 1 (Amy1) genes were in cluster (c). IGF-binding protein 7 (Igfbp7) and solute carrier family 6 (Slc6a6) were genes from cluster (d).

3.2. Ontology Analysis

As shortening events were generally accumulated in gestation and lactation and lengthening events during involution, we performed an ontology analysis using the genes that were detected by APADetect. Shortened genes in gestation (n=561) and lactation (n=389), that have SLR values above 1.5 and lengthened genes in involution (n= 246), that have SLR values below 0.6 were analysed using DAVID annotation tool [102]. The software uses annotations obtained from a range of databases to detect gene clusters that are enriched within the set of genes provided. Genes were classified according to the GOTERM biological processes (Figure 3.5).

Genes involved in processes such as regulation of migration, actin cytoskeleton organization and cell projection (lammelipodium) assembly were enriched within shortened set of genes in gestation. Extensive branching of the ductules and invasion through the fat pads require increased activity of genes involved in cell migration and cytoskeletal organization. Therefore, shortening of the genes associated with these biological processes can possibly result in increased protein amount, which can cause increased activity. Likely, SLR values of genes involved in fatty acid metabolic process were found to be increased during gestation, which can be explained by extensive energy consumption during increased proliferation of epithelia. Shortened genes in lactation, on the other hand, were enriched with genes involved in mammary gland development and lactation processes, that is directly associated with the nature of lactating mammary gland.

Genes involved in immune response and response to wounding were enriched within the lengthened genes during involution. Such genes were previously shown to be upregulated in mRNA level, using the same microarray data, during involution switch after weaning [107]. Therefore the overall increase in mRNA levels might be caused by the increase in long isoform transcription.

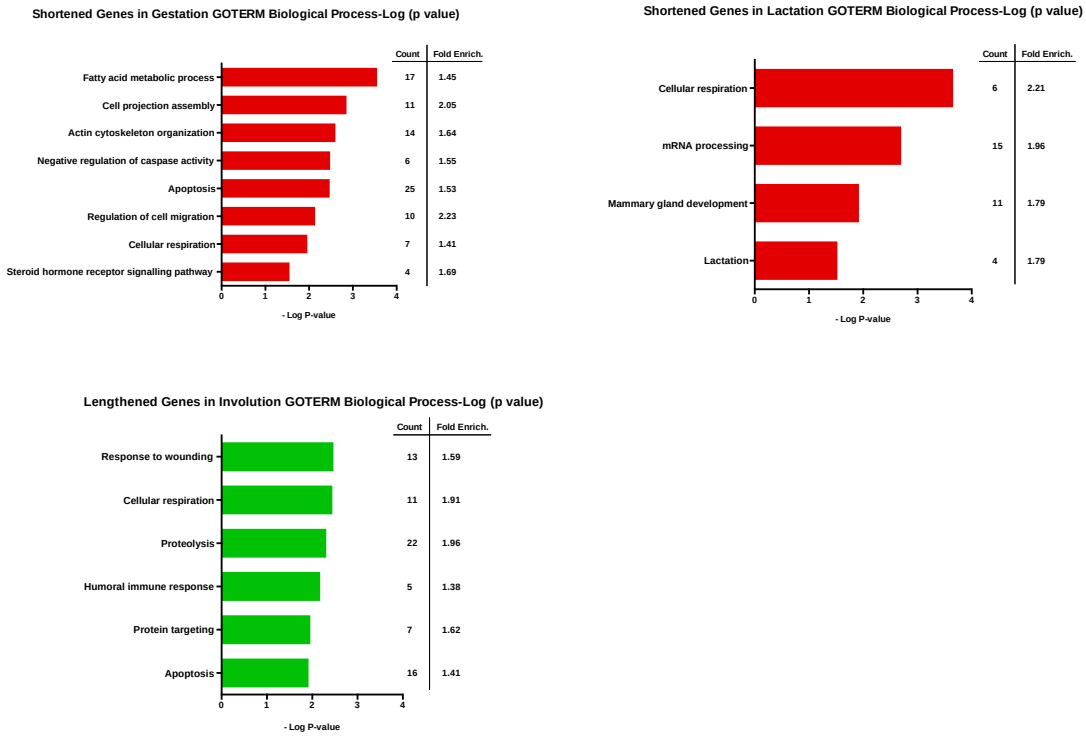

Figure 3.5. Ontology analysis of significant genes.
DAVID ontology analysis of shortened genes (red) in gestation and lactation and lengthened genes (green) during involution based on biological processes. Shortened genes are the genes that had a treated to control SLR ratio above 1.5 and lengthened genes have the ratio below 0.6. Enrichment score represents the fold enrichment value that indicates how that particular biological process activity is abundant in the data set compared to genome. Count refers to the number of genes in the data set that is associated with each GOTERM.

3.3. Experimental Validation of *In Silico* Data

To test the results obtained from APADetect, HC11 mouse mammary epithelial cell line, which was derived from mammary gland of mice in mid-gestation, was used as a model system for gestation and lactation timepoints. These cells are sensitive to lactogenic hormone mix; prolactin, cortisol and insulin. When treated with lactogenic hormones, HC11 cells undergo lactogenic differentiation. Differentiated (treated) HC11 cells start to synthesize the milk protein β-casein and they are organized into alveolar- resembling structures, whereas undifferentiated (untreated) HC11 cells form monolayers with reduced β-casein expression [103, 105].

To optimize treatment conditions of HC11 cells with prolactin, cells were treated with 1μg/ml prolactin for 6, 12, 24, 36, 48 and 72 hours. Then β-casein expression levels were analysed using gene specific primers by quantitative real time PCR (RT-qPCR) (Figure 3.6). Gene expression was normalized to β-casein expression in untreated HC11 cells. Relative β-casein expression was shown to increase gradually starting from 6 hours of treatment until 36 hours and then it decreased. Therefore, 36 hours of treatment was selected for further experiments.

Next, we also tested the dose effect of prolactin on β-casein expression. HC11 cells were treated with 1 μg/ml, 2 μg/ml, 3 μg/ml, 4 μg/ml and 5 μg/ml prolactin for 36 hours and β-casein expression levels were analysed (Figure 3.7). A linear increase in relative β-casein expression with increasing prolactin concentration was observed.

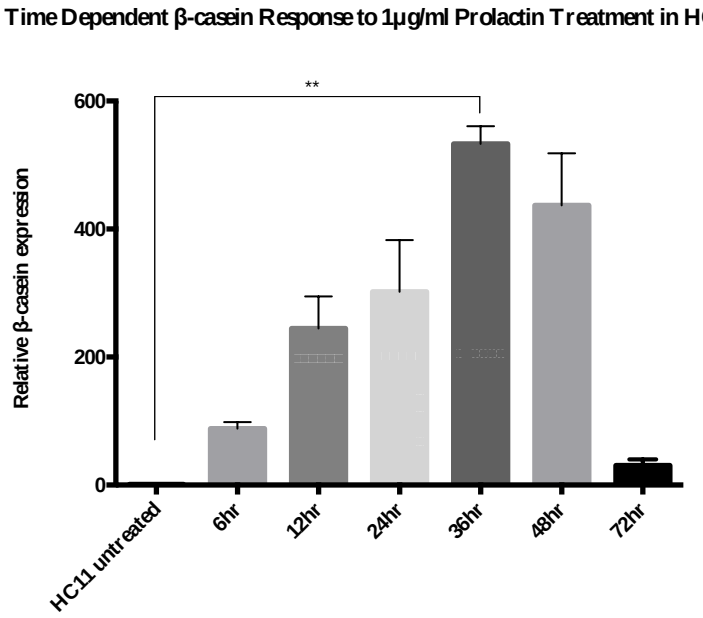

Figure 3.6. Relative quantification of β-casein expression in HC11 cells treated with 1 μg/ml prolactin at six different timepoints.
HC11 cells were treated with 1μg/ml prolactin for 6, 12, 24, 36, 48 and 72 hours. Then RT-qPCR analysis carried out to observe the changes in β-casein expression compared to untreated HC11 cells. Fold change was normalized against the reference gene CycB. Reaction efficiency correction and ΔΔCq method was used for quantification [106]. The baseline for the β-casein expression in untreated HC11 samples was set to 1. Experiment was repeated 2 times with 3 technical replicates. Relative expression of β-casein peaks at 36hr treatment. ** indicates significant difference between 36hr prolactin-treated HC11 cells and untreated HC11 cells, $p<0.01$ (Unpaired t-test with equal SD).

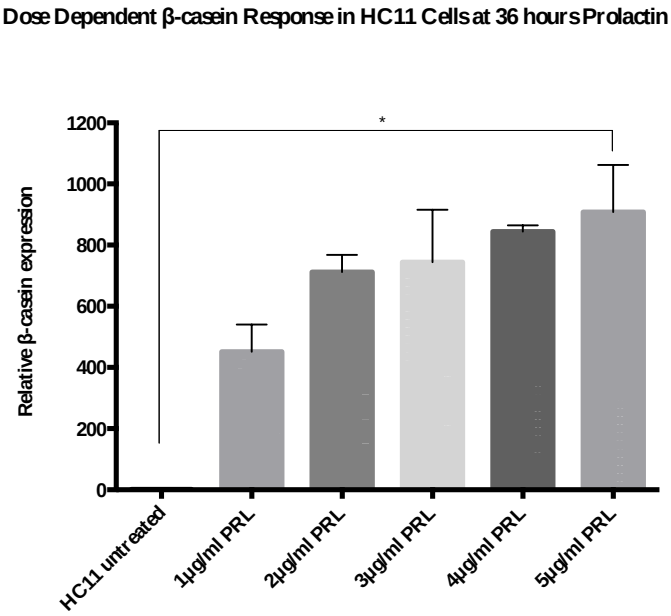

Figure 3.7. Relative quantification of β-casein expression in HC11 cells when treated with different prolactin concentrations for 36 hours.
HC11 cells were treated with 1, 2, 3, 4 and 5µg/ml prolactin for 36 hours. Then RT-qPCR analysis carried out to observe the changes in β-casein expression compared to untreated HC11 cells. Fold change was normalized against the reference gene CycB. Reaction efficiency correction and ΔΔCq method was used for quantification [106]. The baseline for the β-casein expression in untreated HC11 samples was set to 1. Experiment was repeated 2 times with 3 technical replicates. β-casein expression gradually increases and peaks when the cells were treated with 5µg/ml prolactin. * indicates significant difference between 5µg/ml prolactin-treated HC11 cells and untreated HC11 cells, $p<0.05$ (Unpaired t-test with equal SD).

Therefore to obtain maximum effect of prolactin at mRNA expression levels, treatments were optimized at 5 µg/ml prolactin for 36 hours for the rest of the experimental procedure (Figure 3.8).

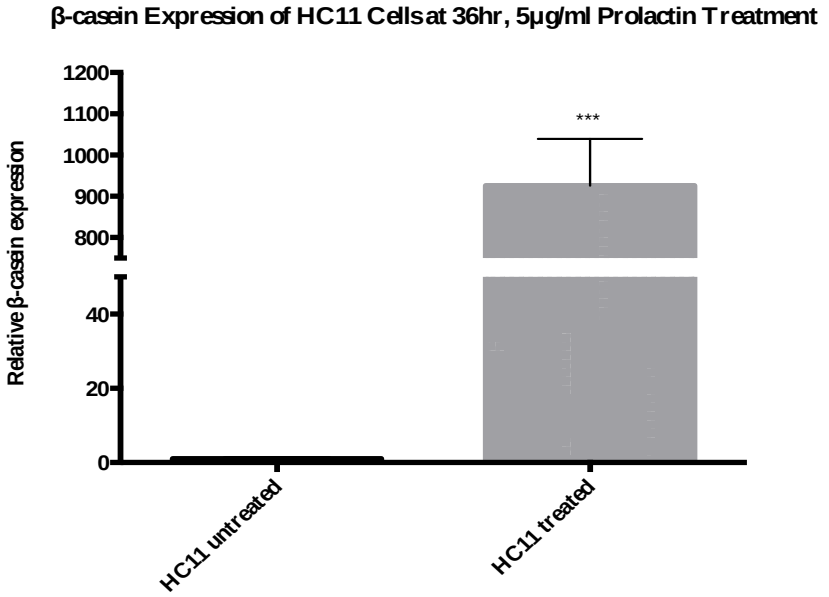

Figure 3.8. Relative quantification of β-casein expression in HC11 cells treated with 5 µg/ml prolactin for 36 hours.
HC11 cells were treated with 5µg/ml prolactin for 36 hours and to confirm lactogenic differentiation, β-casein expression was analyzed by RT-qPCR analysis. Fold change was normalized against the reference gene CycB and relative expression of the gene in untreated HC11 cells were set to 1. Reaction efficiency correction and ΔΔCq method was used for quantification [106]. *** indicates significant difference between prolactin-treated HC11 cells an untreated cells' β-casein expression, $p<0.001$ (Unpaired t-test with equal SD).

3.3.1. Confirmation of Candidate APA Events

To confirm the 3'UTR shortening and lengthening events detected by APADetect *in vitro*, gene specific primers for short and long isoforms of selected genes were designed. All RT-qPCR analyses were done using HC11 untreated cells as gestation model, prolactin-treated HC11 cells as lactation model and 4T1 cells as triple negative breast cancer model. Relative expression levels in prolactin-treated HC11 cells and 4T1 cells were normalized to that of untreated HC11 cells. All experiments were repeated 3 times with 3 technical replicates, unless otherwise noted.

3.3.1.a. Xanthine dehydrogenase (Xdh) Expression

APADetect results indicated significant shortening of Xdh transcripts at 0 day and 5 day lactation timepoints when compared to the virgin timepoint (Figure 3.9). The increase in SLR ratios in lactation timepoints were also significant when compared to 15 day gestation timepoint (mid-gestation). Shortening of the gene was recovered in mammary glands after weaning and because of this trend the gene belonged to cluster b in the heatmap.

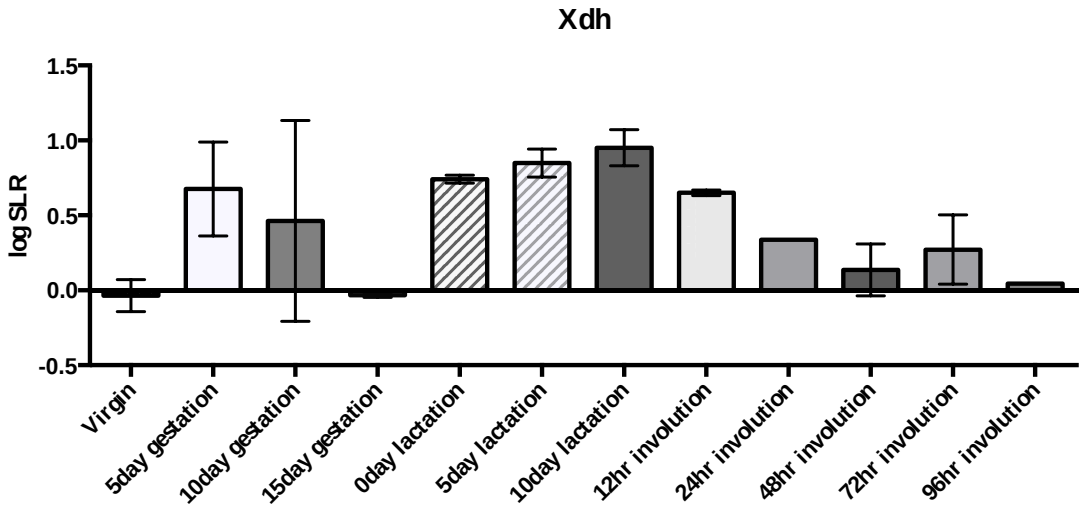

Figure 3.9. Log SLR values of Xdh gene obtained from APADetect at 12 timepoints.
A significant shortening at 0 day and 5 day lactation timepoints (shaded) was observed when compared to virgin the virgin timepoint. Log SLR values drop back to the levels similar to virgin timepoint during involution.

To confirm the shortening events during lactation, short and long isoforms were quantified by RT-qPCR analyses (Figure 3.10). SLR values were calculated by dividing relative expression of short isoforms to that of long isoforms. An increase in both short and long isoforms of the gene in prolactin-treated cells was observed, however the increase in relative expression of short isoform was higher resulting in a high SLR value. A significant increase in the SLR ratio in prolactin-treated HC11 cells confirmed shortening of the gene during lactation. 4T1 cells, on the other hand, have

relative expression values similar to that of untreated HC11 cells. No significant change in SLR value was observed in 4T1 cells.

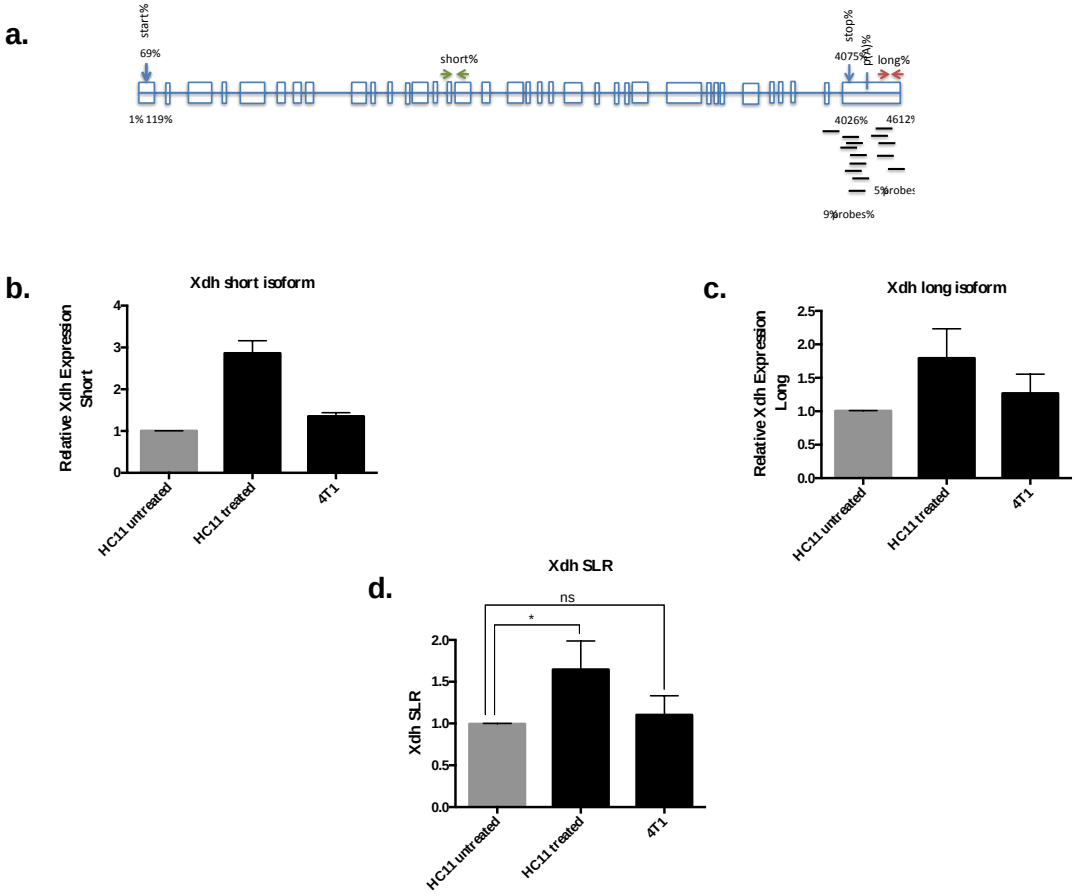

Figure 3.10. Relative quantification of Xdh short and long isoforms in untreated, prolactin-treated HC11 cells and 4T1 cells.
(a) Schematic representation of Xdh gene. Exons are represented as boxes and introns are shown as lines. Locations of proximal (n=9) and distal probes (n=5) and locations of short and long primers are shown. (b) Relative expression of the short isoform of Xdh gene. (c) Relative expression of long isoform of Xdh gene. (d) SLR values obtained by dividing relative expression of short isoform to that of long isoform. The baseline for gene expression in untreated HC11 samples was set to 1 and reaction efficiency correction and ΔΔCq method was used for quantification [106]. One-way ANOVA with Dunnett's multiple comparison post test was performed in GraphPad Prism software (California, USA). * indicates a significant difference between prolactin-treated HC11 cells and untreated HC11 cells ($p<0.05$).

Xdh has long been known as the key enzyme in the catabolism of purines. It oxidises hypoxanthine into xanthine and then xanthine into uric acid [108]. The encoded protein exists in two enzymatic forms, either as oxidase or dehydrogenase. Xanthine

dehydrogenase can be converted to xanthine oxidase by proteolytic modification irreversibly or reversibly by sulfhydryl oxidation [109]. In addition to its role in purine degredation, Xdh is also involved in lipid droplet formation during milk synthesis in lactating mammary gland. It was shown that prolactin induces Xdh expression in mouse mammary gland and mice heterozygous for a loss-of-function mutation in the Xdh gene are unable to maintain lactation [110, 111]. Supporting this, its mRNA expression was found to be upregulated in lactating bovine mammary gland [112]. Immunohistochemical analysis of the Xdh protein revealed its presence in the cytoplasm of epithelial cells that line the terminal ducts in normal breast tissue and enhanced presence in the alveolar epithelium of the lactating tissue. In contrast, no detectable Xdh was present in invasive breast carcinomas [113].

Independent studies have demonstrated the down regulation in Xdh expression levels and this down regulation might be considered as a feature of aggressive breast carcinogenesis. Loss of protein was shown to result in a significant decrease in disease free survival of breast cancer patients [114]. Therefore, Xdh can be considered as a tumour-suppressor gene.

Shortening of the gene can possibly cause an increase in its protein levels in prolactin-treated cells, which can be explained by the fact that the gene is involved in milk protein synthesis.

3.3.1.b. Syndecan 2 (Sdc2) Expression

Array data indicated significant shortening of Sdc2 mRNA during 5 day gestation, 5 day lactation and 12 hour involution timepoints compared to the virgin (Figure 3.11). However, shortening in 5 day lactation is also significant when compared to 15 day gestation timepoint (mid-gestation).

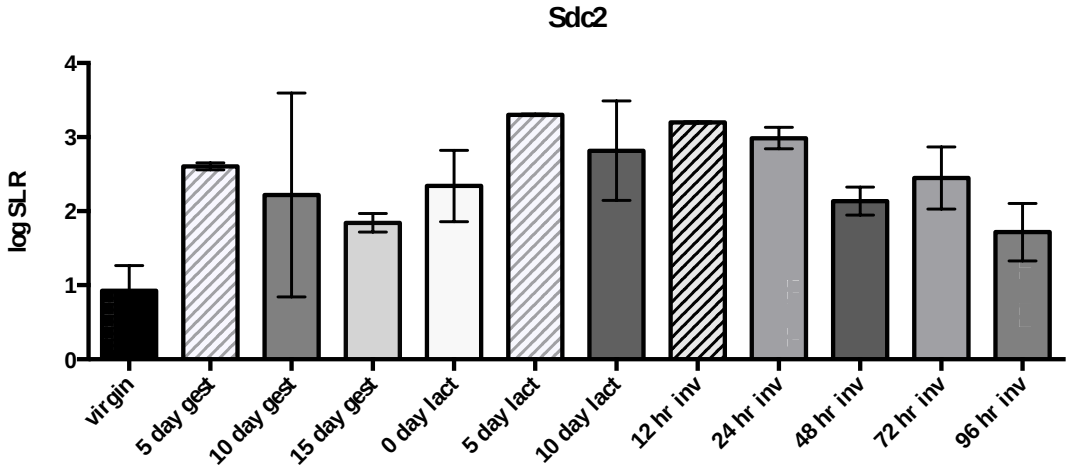

Figure 3.11. Log SLR values of Sdc2 gene obtained from APADetect at 12 timepoints. Thrsp mRNA was found to be significantly shortened at 5 day gestation, 5 day lactation and 12 hour involution timepoints (shaded) when compared to virgin.

To confirm the array data, RT-qPCR analyses were performed using primers specific for short and long isoforms of the gene (Figure 3.12 a,b,c). Then SLR values were calculated by dividing relative expression of short isoforms to that of long isoforms (Figure 3.12 d).

Sdc2 is a cell surface transmembrane heparan sulfate (HS) proteoglycan. It is known to be involved in cell signalling, cell binding and cytoskeletal organization together with other members of the syndecan family [115]. Expression of the gene was found to be upregulated in pancreatic cancer, melanoma, colon cancer and in fibrosarcomas. Increased protein levels are thought to enhance cell adhesion, proliferation and migration in cancer cells [116, 117].

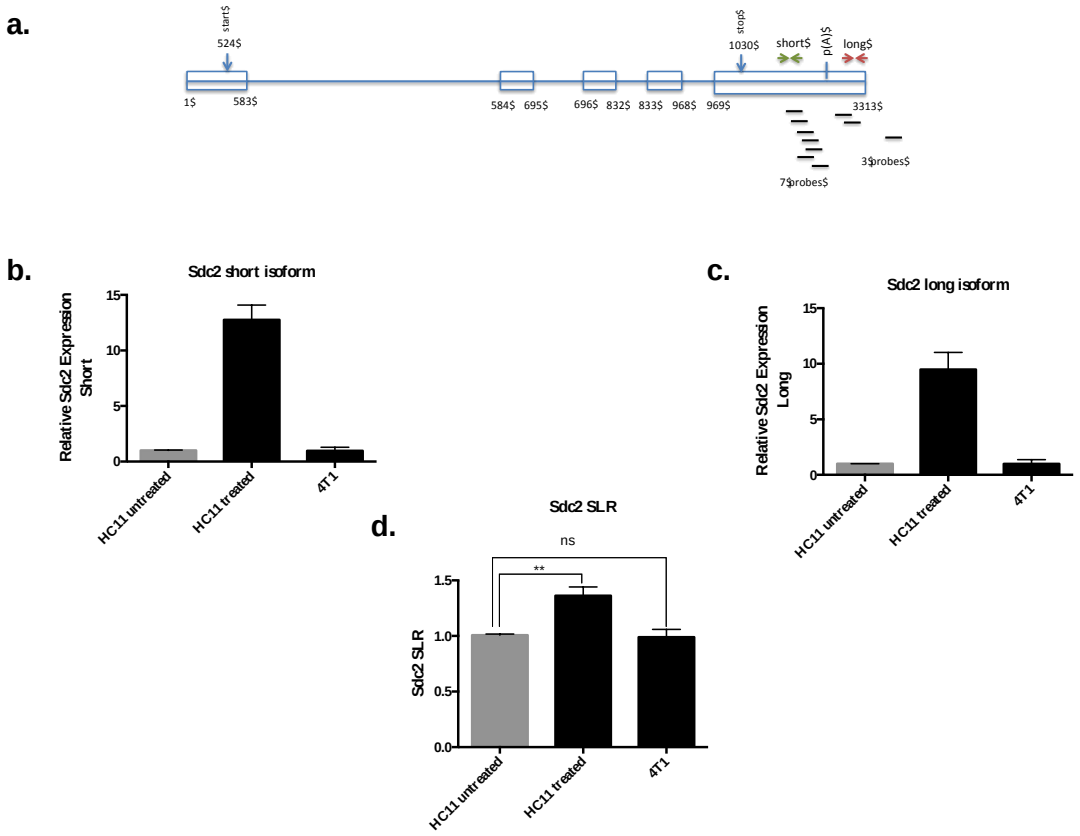

Figure 3.12. Relative quantification of Sdc2 short and long isoforms in untreated HC11, prolactin-treated HC11 and 4T1 cells.
(a) Schematic representation of Sdc2 gene. Exons are represented as boxes and introns are shown as lines. Locations of proximal (n=7) and distal probes (n=3) and locations of short and long primers are shown. (b) Relative expression of the short isoform of Sdc2 gene. (c) Relative expression of long isoform of Sdc2 gene. (d) SLR values obtained by dividing relative expression of short isoform to that of long isoform. The baseline for gene expression in untreated HC11 samples was set to 1 and reaction efficiency correction and ΔΔCq method was used for quantification [106]. For significance analysis one-way ANOVA with Dunnett's multiple comparison post test was performed in GraphPad Prism software (California, USA). ** indicates a significant difference between prolactin-treated HC11 cells and untreated HC11 cells ($p<0.01$).

Sdc2 is normally expressed in highly invasive TNBC human breast cell line MDA-MB231. Invasion ability and degradation of type I collagen in the ECM were inhibited when Sdc2 was depleted in these cells. Additionally, reduction of cell invasiveness was also accompanied by changes in cell-cell adhesion and actin cytoskeletal organization, where focal adhesions, microfilament bundles and cadherin-11 containing adherens junctions were enhanced [116]. Immunohistochemical analysis performed by staining tissue sections of breast cancer patients and healthy women

revealed cytoplasmic/cell membrane localisation of the protein in normal tissue and overexpression with some nuclear localisation in tumours [116].

In another study, where Sdc2 was overexpressed in an osteosarcoma cell line that normally have low expression, induction of apoptosis was observed [118].

3.3.1.c. Thyroid Hormone Responsive Gene (Thrsp)

According to the APADetect results, Thrsp was shown to be significantly shortened during 5 day gestation, 0 day and 10 day lactation when compared to the virgin timepoint (Figure 3.13). Shortening of the transcript during lactation was also significant when compared to gestation 15 day timepoint (mid-gestation).

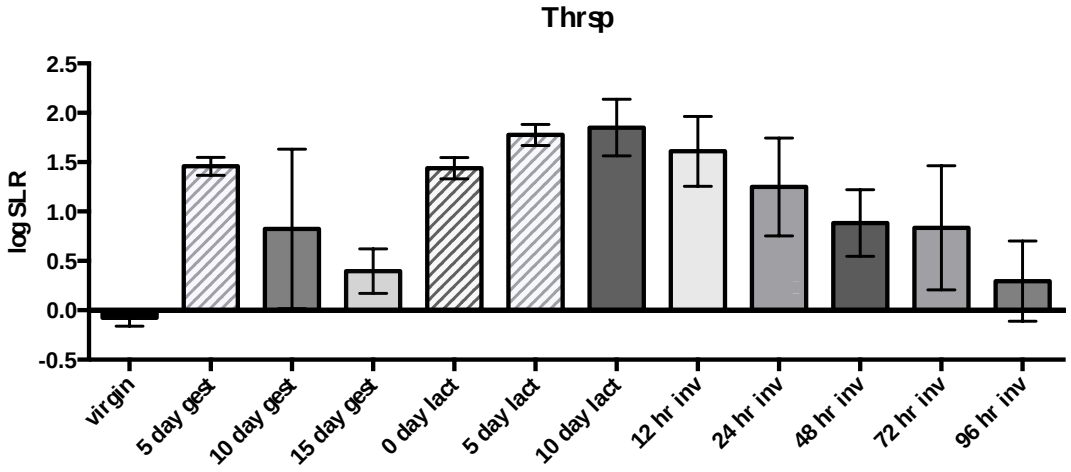

Figure 3.13. Log SLR values of Thrsp gene obtained from APADetect at 12 timepoints. Thrsp mRNA was found to be significantly shortened at 5 day gestation, 0 day and 5 day lactation timepoints (shaded) when compared to virgin. During involution, recovery of the shortening was observed.

To confirm shortening of the transcript during lactation, RT-qPCR analysis were done using short and long isoform specific primers (Figure 3.14). SLR values were calculated by dividing relative expression levels of short isoforms to that of long isoforms. An overall increase in relative expression levels of Thrsp short and long isoforms were observed in prolactin-treated HC11 cells and 4T1 cells. SLR value was

significantly increased in prolactin-treated HC11 cells but no significant change was observed in 4T1 cells when compared to untreated HC11 cells.

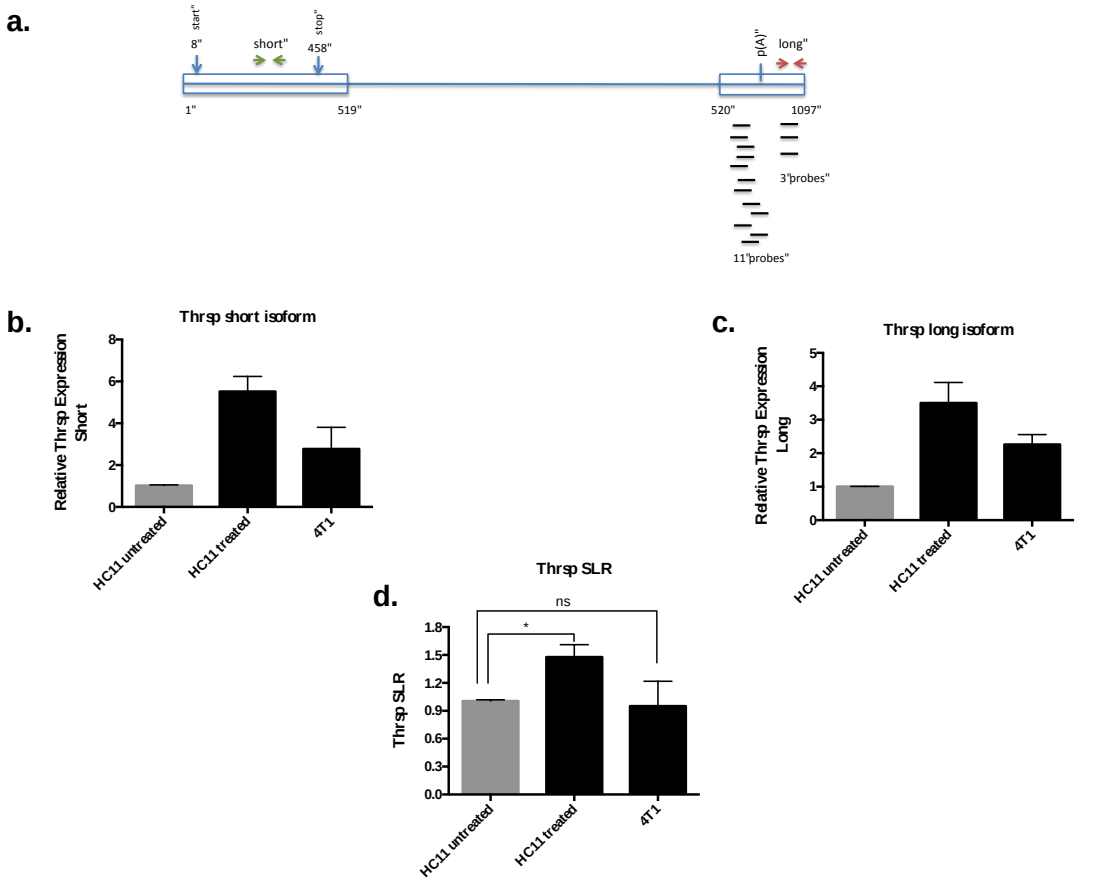

Figure 3.14. Relative quantification of Thrsp short and long isoforms in untreated HC11, prolactin-treated HC11 and 4T1 cells.
(a) Schematic representation of Thrsp gene. Exons are represented as boxes and introns are shown as lines. Locations of proximal (n=11) and distal probes (n=3) and locations of short and long primers are shown. Reaction efficiency correction and ΔΔCq method was used for quantification. (b) Relative expression of the short isoform of Thrsp gene. (c) Relative expression of long isoform of Thrsp gene. (d) SLR values obtained by dividing relative expression of short isoform to that of long isoform. The baseline for gene expression in untreated HC11 samples was set to 1 and reaction efficiency correction and ΔΔCq method was used for quantification [106]. One-way ANOVA with Dunnett's multiple comparison post test was performed in GraphPad Prism software (California, USA). * indicates a significant difference between prolactin-treated HC11 cells and untreated HC11 cells ($p<0.05$).

Thrsp, which is also known as Spot14 or S14, is a crucial protein for cellular de novo lipogenesis. Thrsp was shown to be upregulated in lactating tissue when compared

with the dry period in goat mammary tissue [119]. Its role in the regulation of mammary lipid synthesis during milk production was also confirmed in bovine breast [120]. In addition to its role during milk fat synthesis, Thrsp was found to be an important regulator of epithelial proliferation in the mammary gland as well. Therefore shortening of the gene, that might cause an increased protein level, was indeed expected in prolactin-treated HC11 cells.

The gene was also shown to be strongly expressed especially in lipogenic tumours [121]. In lipogenic tumours the activity of the fatty acid synthase (FASN) enzyme is upregulated, conferring growth and survival advantages rather than functioning as an anabolic energy-storage pathway in tumours [122]. Overexpression of Thrsp was shown to upregulate the expression of some enzymes involving FASN in goat mammary epithelial cells [119].

3.3.1.d. IGF-Binding Protein 7 (Igfbp7)

Igfbp7 was found to be significantly lengthened during 10 day lactation, 12, 24 and 48 hour involution timepoints in APADetect (Figure 3.15). Lengthening was also significant at 10 day lactation when compared to 15 day gestation timepoint.

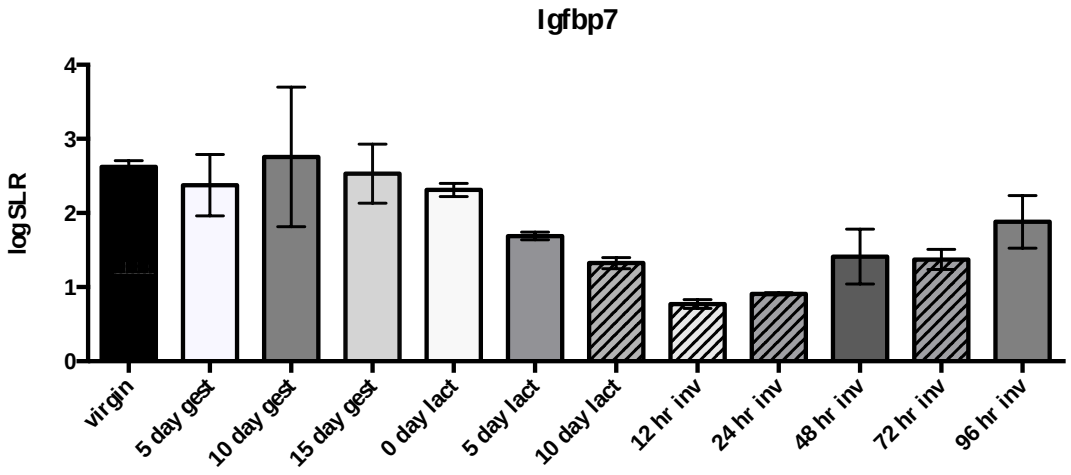

Figure 3.15. Log SLR values of Igfbp7 gene obtained from APADetect at 12 timepoints. Igfbp7 mRNA was found to be significantly lengthened at 10 day lactation, 12 hour, 24 hour and 72 hour involution timepoints (shaded) when compared to virgin. Recovery of early lengthening events was observed at 96 hour timepoint.

To confirm the lengthening event in lactation, RT-qPCR analyses were done using primers specific for short and long isoforms of Igfbp7 gene (Figure 3.16). SLR values were then calculated as explained before. Relative expressions of short and long isoforms were around 7-fold increased in 4T1 cells. However, SLR ratio was not significantly changed in 4T1 cells. In prolactin-treated HC11 cells, on the other hand, a significant lengthening of Igfbp7 was observed compared to untreated HC11 cells.

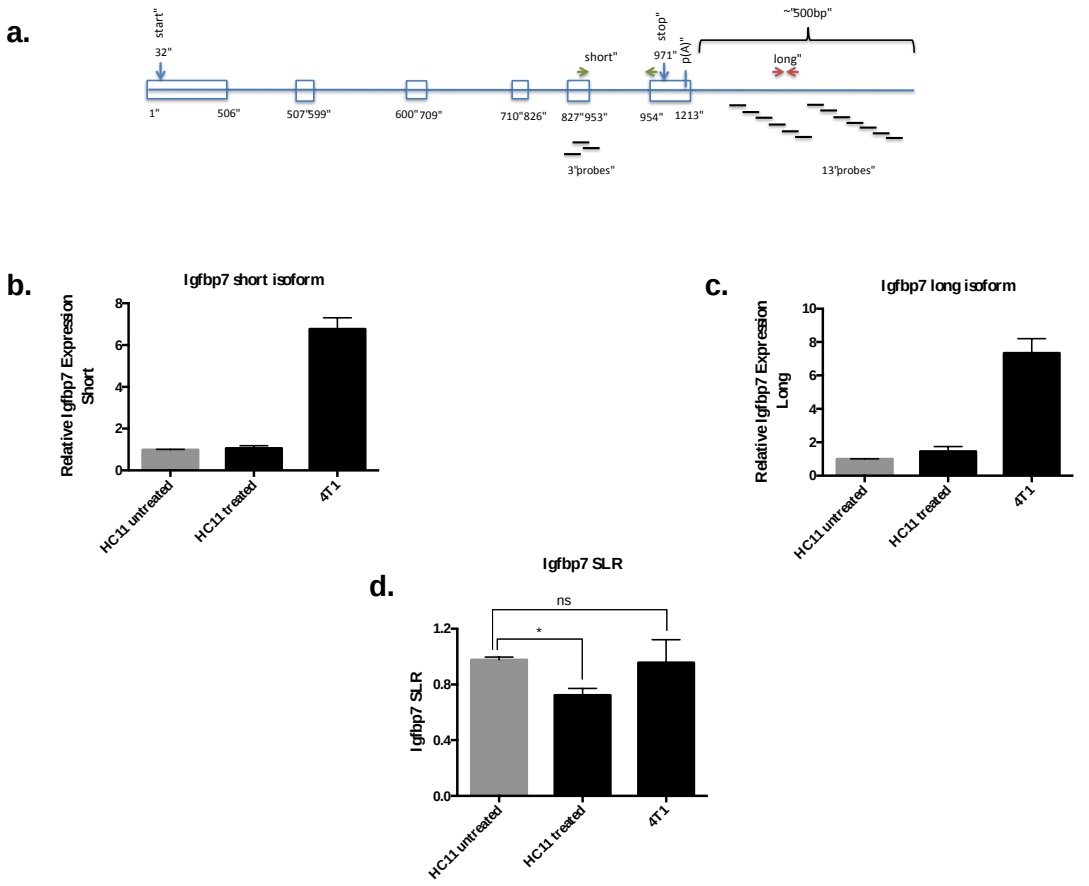

Figure 3.16. Relative quantification of Igfbp7 short and long isoforms in untreated, prolactin-treated HC11 cells and 4T1 cells.
(a) Schematic image of Igfbp7 gene. Exons are represented as boxes and introns as lines. Locations of proximal (n=3) and distal probes (n=13) and locations of short and long primers are shown. (b) Relative expression of the short isoform of Igfbp7 gene. (c) Relative expression of long isoform of Igfbp7 gene. (d) SLR values obtained by dividing relative expression of short isoform to that of long isoform. The baseline for gene expression in untreated HC11 samples was set to 1 and reaction efficiency correction and ΔΔCq method was used for quantification [106]. For significance analysis one-way ANOVA with Dunnett's multiple comparison post test was performed in GraphPad Prism software (California, USA). * indicates a significant difference between prolactin-treated HC11 cells and untreated HC11 cells ($p<0.05$).

Igfbp7 is one of the known sixteen Igfbp superfamily proteins and it is also named as Igf-binding protein related protein 1 (Igfbp-rp1) [123]. It binds to Igf1, Igf2 and insulin but have much lower affinity than other Igfbps. This suggests that Igfbp7 might have different functions from other Igf-binding proteins [63]. It was previously reported that loss of Igfbp7 cause retarded mammary gland development in virgin mice and pregnant Igfbp7-null mice develop glands with reduced numbers of alveolar structures. Most significantly, loss of Igfbp7 induces precocious involution in lactating mammary gland through reduced Stat5 and Akt signaling [63].

In breast cancers it is thought that Igfbp7 might act as a tumor suppressor gene. Supporting this, immunohistochemical analysis revealed a downregulation of Igfbp7 in half of the human metastatic breast tumors tested. Also overexpression of the gene in TNBC cell lines, which normally have low or no expression, results in induction of apoptosis [124, 125]. It was shown that in tumor cells activated oncogenes such as BRAF can induce apoptosis by triggering oncogene-activated senescence response through Igfbp7 [126].

3.3.1.e. Heterogeneous nuclear ribonucleoprotein A1 (Hnrnpa1)

Hnrnpa1 mRNA was shown to be significantly lengthened during 10 day lactation and 12 hour involution timepoints (Figure 3.17). Lengthening at 10 day lactation is also significant when compared to 15 day gestation (mid-gestation) timepoint.

Primers were designed for short and long isoforms of the Hnrnpa1 transcripts and RT-qPCR analyses were done to confirm the lengthening event during lactation. SLR values were calculated by dividing relative expression of the short isoform to that of long isoform (Figure 3.18). SLR values indicated a significant lengthening in prolactin-treated HC11 cells when compared to untreated HC11 cells.

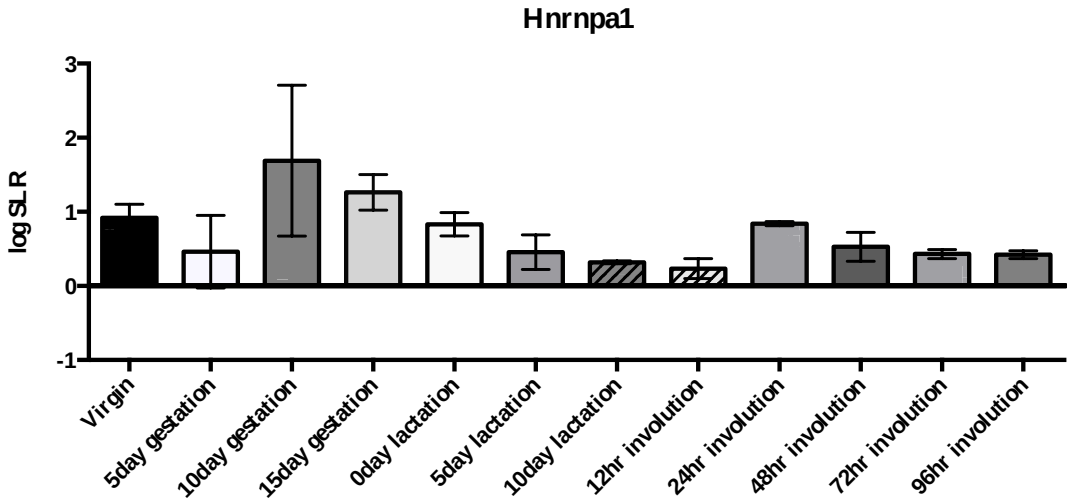

Figure 3.17. Log SLR values of Hnrnpa1 gene obtained from APADetect at 12 timepoints.
Hnrnpa1 mRNA was found to be significantly lengthened at 10 day lactation and 12 hour involution timepoints (shaded) when compared to virgin.

Hnrnpa1 is an RNA binding protein, which is a member of the A/B superfamily of hnRNPs. HnRNP proteins associate with pre-mRNAs in the nucleus and known to have a variety of functions in gene mRNA transport, mRNA stability, miRNA maturation, splicing and telomere biogenesis [127, 128]. Hnrnpa1 is one of the most abundant core proteins of hnRNP complex and is primarily involved in gene splicing [129].

Hnrnpa1 was also found to be closely related in tumorigenesis and is deregulated in a variety of tumours including oral squamous cell carcinoma, colorectal cancers and in breast cancers [130-132]. Overexpression was found to promote tumour invasion and to be connected to poor prognosis in hepatocellular carcinoma (HCC) [127]. Therefore it was considered as a potential oncogene. Supporting this, knock-down of the gene in cancer cells resulted in induction of G2/M arrest, apoptosis, reduced cell viability and impaired invasion capability in breast cancer cell lines [131-133].

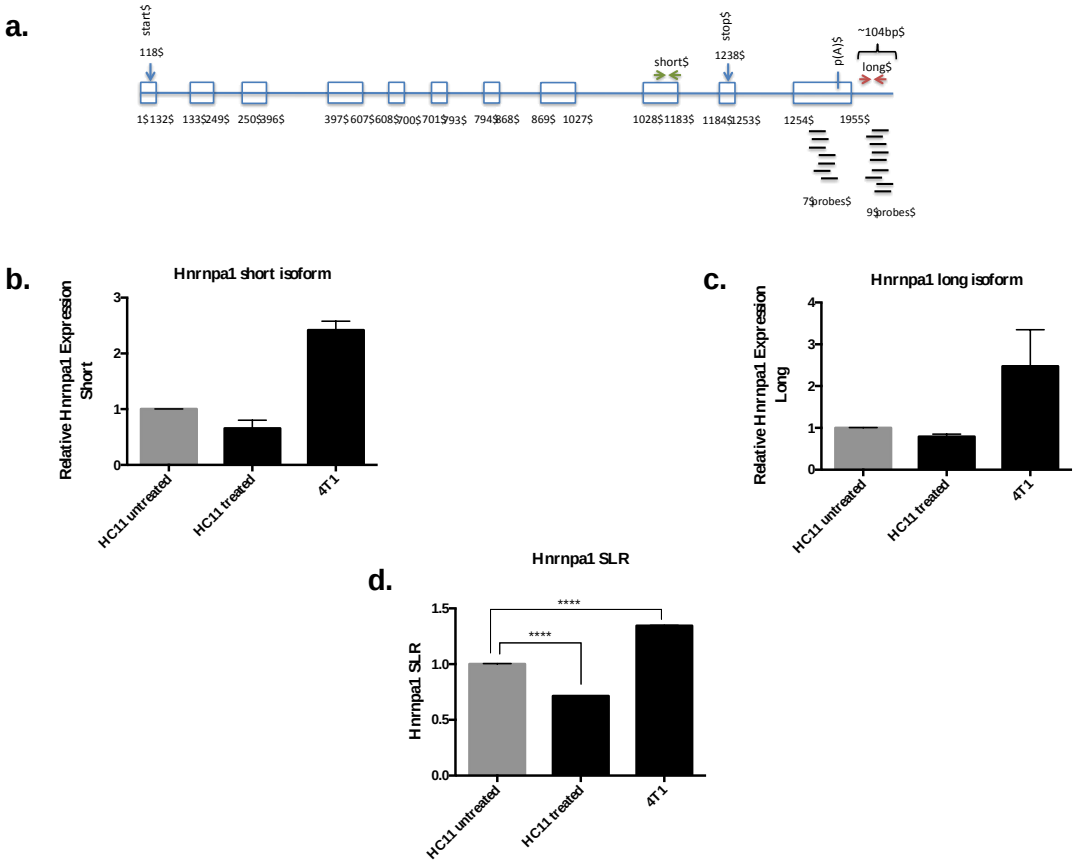

Figure 3.18. Relative quantification of Hnrnpa1 short and long isoforms in untreated, prolactin-treated HC11 cells and 4T1 cells.
(a) Schematic representation of Hnrnpa1 gene. Exons are represented as boxes and introns are shown as lines. Locations of proximal (n=7) and distal probes (n=9) and locations of short and long primers are shown. (b) Relative expression of the short isoform of Hnrnpa1 gene. (c) Relative expression of long isoform of Hnrnpa1 gene. (d) SLR values obtained by dividing relative expression of short isoform to that of long isoform. The baseline for gene expression in untreated HC11 samples was set to 1 and reaction efficiency correction and ΔΔCq method was used for quantification [106]. For significance analysis one-way ANOVA with Dunnett's multiple comparison post test was performed in GraphPad Prism software (California, USA). **** indicates a significant difference ($p<0.0001$).

Hnrnpa1 is involved in alternative splicing of numerous important genes. For example, in a study using breast cancer cell lines, it was shown that Hnrnpa1 regulates splicing of CD44 gene, which is a transmembrane receptor for hyaluronic acid and is also a breast cancer stem cell marker protein. When Hnrnpa1 was knocked down from breast cancer cell lines, overall expression level of CD44 was not changed but the amount of different splice variants were significantly altered [132, 134]. In addition to its role in

cancer stem cells, CD44 is also functionally required in the lactating mammary gland [135].

In another study, Hnrnpa1 was shown to regulate alternative splicing of the pyruvate kinase enzyme, which the key enzyme involved in aerobic glycolysis [136]. Aerobic glycolysis is the phenomenon that the tumour cells consuming glucose very avidly and converting them to lactate to obtain the highest possible energy. In human gliomas, Hnrnpa1 was shown to be upregulated by the oncogenic transcription factor c-Myc and produce the alternative splice variant of the pyruvate kinase that promote aerobic glycolysis [136].

Therefore, changing protein levels due to alternative polyadenylation of the gene may cause alternative splicing of such key genes that are involved in gland development and tumorigenesis.

For further analysis we investigated the protein expression levels of the Hnrnpa1 in the development model and 4T1 cells (Figure 3.19).

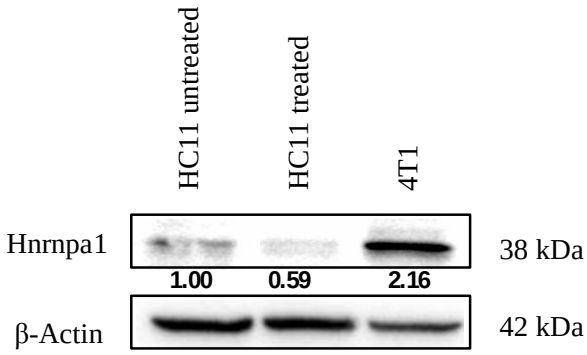

Figure 3.19. Protein expression levels of Hnrnpa1 in untreated HC11 cells, prolactin-treated HC11 cells and 4T1 cells.
Western blot analysis of Hnrnpa1 expression in untreated, prolactin-treated HC11 cells and 4T1 cells. 70 µl of total cell lysates from cells were used to quantify Hnrnpa1 protein levels. Beta-actin was used as a loading control.

Results indicated a significant decrease in the protein levels of Hnrnpa1 in prolactin-treated HC11 cells when compared to untreated cells and a significant increase in 4T1 cells. This partly confirms the effect of lengthening and shortening events in 3'UTR region of Hnrnpa1 transcripts on protein levels. Significant lengthening of the transcript seems to result in decreased protein production and significant shortening results in significantly increased protein production within the cells.

To further investigate the role of Hnrnpa1 protein in breast cancers we aim to investigate the APA events and protein expression levels in human breast cell lines. If any significant changes are observed further functional analyses will be carried out to understand the exact role of the protein in development and breast tumorigenesis.

3.4. Cstf2 and CFIm25 Expression Analysis

To have an insight whether the APA-dependent changes in transcript lengths are caused by changes in the expression levels of proteins involved in the APA complex, we investigated the mRNA and protein expression levels of Cstf2 and CFIm25 genes in untreated and prolactin-treated HC11 cells and in 4T1 cells.

Previous work done in our lab has shown that especially Cstf2 mRNA and protein levels change dramatically between normal breast and breast cancer cell lines, as well as in response to estrogen (E2) or epidermal growth factor (EGF) treatments to cell lines [93, 137]. When Cstf2 protein levels are elevated within a cell, selection of the proximal poly (A) sites are favoured, resulting in shortening of the transcripts [99].

Therefore to analyse whether Cstf2 mRNA and protein expression changes in prolactin-treated HC11 cells and in 4T1 cells we carried out RT-qPCR analysis and western blot analysis and compared expression levels with untreated HC11 cells (Figure 3.20a). Supporting the decrease in significantly shortened gene numbers in lactation compared to gestation timepoints (Figure 3.5), we observed a significant decrease in both mRNA and protein levels of Cstf2 in prolactin-treated HC11 cells.

4T1 cells, on the other hand, have significantly elevated mRNA and protein levels (Figure 3.20b).

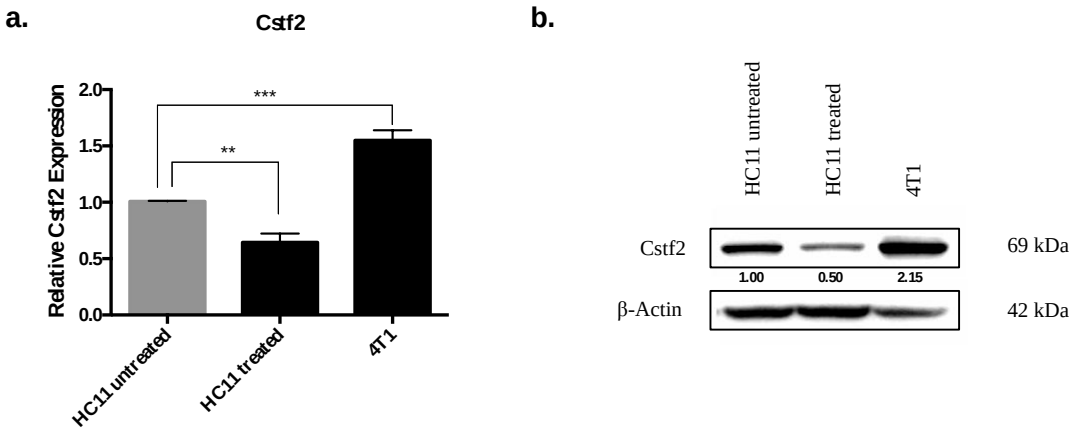

Figure 3.20. mRNA and protein expression levels of Cstf2 in untreated HC11 cells, prolactin-treated HC11 cells and 4T1 cells.
(a) Relative mRNA expression of Cstf2 was quantified by RT-qPCR analyses. The baseline for gene expression in untreated HC11 samples was set to 1 and reaction efficiency correction and ΔΔCq method was used for quantification [106]. For statistical analysis one-way ANOVA with Dunnett's multiple comparison post test was performed using GraphPad Prism (California, USA) software. ** ($p<0.01$) and *** ($p<0.001$) indicates statistical significance. (b) Western blot analysis of Cstf2 expression in untreated, prolactin-treated HC11 cells and 4T1 cells. 70 µl of total cell lysates from cells were used to quantify Cstf2 protein levels. Beta-actin was used as a loading control.

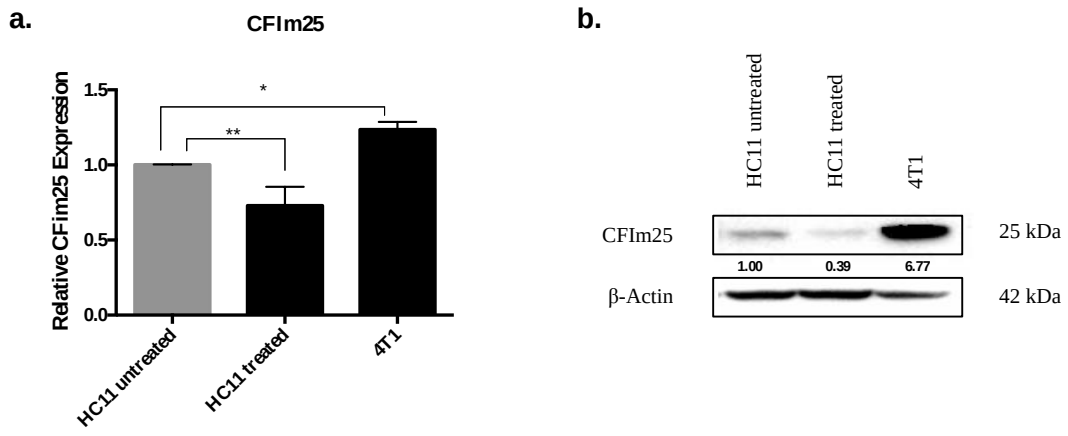

Figure 3.21. mRNA and protein expression levels of CFIm25 in untreated HC11 cells, prolactin-treated HC11 cells and 4T1 cells.
(a) Relative mRNA expression of CFIm25 was quantified by RT-qPCR analyses. The baseline for gene expression in untreated HC11 samples was set to 1 and reaction efficiency correction and ΔΔCq method was used for quantification [106]. For statistical analysis one-way ANOVA with Dunnett's multiple comparison post test was performed using GraphPad Prism (California, USA) software. ** ($p<0.01$) and *** ($p<0.001$) indicates statistical significance. (b) Western blot analysis of CFIm25 expression in untreated, prolactin-treated HC11 cells and 4T1 cells. 70 µl of total cell lysates from cells were used to quantify CFIm25 protein levels. Beta-actin was used as a loading control.

CFIm25 is another protein involved within the polyadenylation complex. RT-qPCR and western blot analyses revealed that the changes in its mRNA and protein levels were the same as the changes in Cstf2 levels in HC11 and 4T1 cells (Figure 3.21). CFIm25 levels are significantly decreased in prolactin-treated cells and significantly increased in 4T1 cells.

Previous work on the CFIm25 gene in glioblastoma cells suggested that an increased level of the protein might lead to the selection of distal poly (A) sites, resulting in lengthening of mRNAs and when depleted it causes increased cell proliferation [138]. However, our data contradicts with this finding as CFIm25 levels were found to be elevated in highly proliferative 4T1 cells and untreated HC11 cells. This might be a result of differentially regulated cell types in different tissues and needs further experimental confirmation.

CHAPTER 4

CONCLUSION

Alternative polyadenylation (APA) has emerged as an important post-transcriptional regulation of gene expression in recent years [99]. As a result of the use of alternative poly (A) sites in the 3'UTRs of pre-mRNAs, transcripts of different lengths can be produced from a single gene. Shortened or lengthened transcripts will eventually have less or more binding sites respectively for regulators such as RNA-binding proteins (RBPs) and microRNAs (miRNAs). As a result, mRNAs with different lengths may differ in their stability, turnover ability, subcellular location etc [86].

To date it was shown that APA is involved in different physiological conditions and cellular contexts including metabolic regulation, stem cell differentiation, white blood cell activation and in complex diseases such as cancer [93]. In cancers, specifically, it was shown that there is a tendency towards proximal poly (A) site usage [95]. In general, shorter 3'UTRs result in increased stability of the mature mRNA as they are targeted by less miRNAs, which in turn causes an increased protein level. Shortening in the 3' UTR sites of oncogenes might result in activation of these proteins. Further investigation revealed that the tendency towards shortening is also abundant in highly proliferative cells, independent of their health status [94].

In this work we investigated the APA patterns during different stages of mammary gland development. Mammary gland is a dynamic organ, changing dramatically in response to circulating hormone levels to meet the infants' needs during gestation, lactation and involution periods in women. It has long been an area of interest whether pregnancy-lactation cycle increases the risk of breast tumorigenesis or it acts as a protective mechanism.

Many studies have investigated the similarities and differences between these stages and breast cancer. In general, it is clear that the cellular processes in epithelial cells during gestation is parallel with tumorigenesis by means of the signalling pathways that are upregulated in both and also activated mechanisms such as reactivation of proliferation in differentiated cells and invasion of the ducts through the fat pads. In this context, it is predicted that pregnancy might increase the risk of breast cancer to a certain degree [5, 19]. However, association between lactation period and tumorigenesis is more contradictive and needs more investigation. There are some epidemiological studies, which indicate that breastfeeding decreases the risk of breast cancer in women. However as all these studies were done using data obtained from nulliparous women and women who have gone through a full pregnancy-lactation cycle it is not possible to understand the effect of lactation period alone in tumorigenesis [16].

Therefore to begin understanding the role of lactation and breast cancers we re-analyzed a microarray data, which contained expression information of samples taken from abdominal mammary glands of virgin, pregnant (gestation day 5, 10, 15), lactating (day 0, 5 and 10) and involuting (hours 12, 24, 48, 72 and 96) mice [98]. Data was analyzed with APAdetect program, where all 11 timepoints were compared to the virgin animal data separately. The results indicated an accumulation of mRNA shortening events mostly in gestation, some in lactation but none in involution. Lengthening events, on the other hand, were almost exclusively observed in involution timepoints and very few in late lactation timepoints. Ontology analysis revealed that the genes shortened during gestation were highly related to processes involved in proliferative and invasive cell populations, such as regulation of migration and actin cytoskeleton organization. Also genes shortened during lactation were related to this stage of mammary development as significant clusters included developmental growth and lactation. This supported the idea that APA might somehow be involved in mammary gland development as an important post-transcriptional regulation mechanism.

To confirm *in silico* data we obtained from APADetect and to investigate any similarities between APA regulation during gestation and lactation stages we performed RT-qPCR analyses using untreated HC11 cells as gestation model, prolactin-treated (differentiated) HC11 cells as lactation model and TNBC 4T1 cells. 3 genes that were significantly shortened during lactation (Xdh, Sdc2, Thrsp) and 2 genes significantly lengthened during gestation (Igfbp7 and Hnrnpa1) were selected as a test group. These genes were previously somehow related to tumorigenesis and also have potential roles in breast development. RT-qPCR results on selected genes have confirmed the SLR values detected *in silico*.

In addition to confirming the validity of our APA detection approach and *in silico* results, we have detected 4T1 cells to be more similar to the gestation status modeled by the untreated HC11 cells. For Xdh, Thrsp and Sdc2 genes, the SLR values for the test genes in the untreated HC11 and 4T1 cells were comparably lower than that of prolactin treated cells. In contrast for Igfbp7 and Hnrnpa1 genes the SLR values in untreated HC11 and 4T1 cells were comparably higher than that of prolactin-treated HC11 cells.

Next, to begin understanding the mechanism causing different poly (A) site selections, we investigated the expression levels of two important proteins involved in the APA complex, Cstf2 and CFIm25. A significant decrease in both mRNA and protein levels of both genes in prolactin-treated cells was observed. In contrast, their levels were significantly increased in 4T1 cells, also suggesting a potential protective role of lactation from tumorigenesis. The prolactin hormone may somehow have a role in regulating the expression of APA proteins, which in turn causes the changes in APA patterns. The direct or indirect effect of prolactin on APA machinery proteins will have to be further investigated.

Here, we provide a platform for novel gene discovery for developmentally regulated APA isoforms that may also have roles in breast cancers. Further studies will follow with two main aims; 1) to better understand the mechanism of APA regulation in

mammary gland development and 2) to investigate the APA isoforms that differ in gestation and lactation states for their potential roles as oncogene or tumor suppressors in breast cancers.

While these results will have to be confirmed at a larger group of breast cancer cell lines and patient samples, our results provide preliminary support for the hypothesis that pregnancy may be associated with high risk of breast cancer, whereas lactation may provide protection against tumorigenesis.

REFERENCES

1. Russo, J. and I.H. Russo, *Development of the human breast.* 2004. **49**(1): p. 2–15.
2. Neville, M.C., T.B. McFadden, and I. Forsyth, *Hormonal regulation of mammary differentiation and milk secretion.* J Mammary Gland Biol Neoplasia, 2002. **7**(1): p. 49-66.
3. Javed, A. and A. Lteif, *Development of the Human Breast.* Seminars in Plastic Surgery, 2013. **27**(1): p. 5-12.
4. Russo, J., R. Rivera, and I.H. Russo, *Influence of age and parity on the development of the human breast.* Breast Cancer Res Treat, 1992. **23**(3): p. 211-8.
5. Wiseman, B.S. and Z. Werb, *Stromal effects on mammary gland development and breast cancer.* Science, 2002. **296**(5570): p. 1046-9.
6. Hassiotou, F. and D. Geddes, *Anatomy of the human mammary gland: Current status of knowledge.* Clin Anat, 2013. **26**(1): p. 29-48.
7. Russo, J., et al., *Influence of age and gland topography on cell kinetics of normal human breast tissue.* J Natl Cancer Inst, 1987. **78**(3): p. 413-8.
8. Humphreys, R.C., et al., *Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis.* Development, 1996. **122**(12): p. 4013-22.
9. Malgorzata Gajewska, K.Z.a.T.M., *Autophagy - A Double-Edged Sword - Cell Survival or Death?*, in *Autophagy in Development and Remodelling of Mammary Gland*. 2013, InTech.
10. Pang, W.W. and P.E. Hartmann, *Initiation of human lactation: secretory differentiation and secretory activation.* J Mammary Gland Biol Neoplasia, 2007. **12**(4): p. 211-21.
11. Medina, D., *The mammary gland: a unique organ for the study of development and tumorigenesis.* J Mammary Gland Biol Neoplasia, 1996. **1**(1): p. 5-19.
12. Sternlicht, M.D., et al., *Hormonal and local control of mammary branching morphogenesis.* Differentiation, 2006. **74**(7): p. 365-81.
13. Kochenour, N.K., *Lactation suppression.* Clin Obstet Gynecol, 1980. **23**(4): p. 1045-59.
14. Jindal, S., et al., *Postpartum breast involution reveals regression of secretory lobules mediated by tissue-remodeling.* Breast Cancer Research, 2014. **16**(2): p. 1.
15. Strange, R., et al., *Apoptotic cell death and tissue remodelling during mouse mammary gland involution.* Development, 1992. **115**(1): p. 49-58.
16. Russo, J., et al., *Breast differentiation and its implication in cancer prevention.* Clin Cancer Res, 2005. **11**(2 Pt 2): p. 931s-6s.
17. Kwan, M.L., et al., *Breastfeeding, PAM50 tumor subtype, and breast cancer prognosis and survival.* J Natl Cancer Inst, 2015. **107**(7).
18. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-752.

19. Djonov, V., A.C. Andres, and A. Ziemiecki, *Vascular remodelling during the normal and malignant life cycle of the mammary gland.* Microsc Res Tech, 2001. **52**(2): p. 182-9.
20. Fata, J.E., Z. Werb, and M.J. Bissell, *Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes.* Breast Cancer Res, 2004. **6**(1): p. 1-11.
21. Bussard, K.M. and G.H. Smith, *The mammary gland microenvironment directs progenitor cell fate in vivo.* Int J Cell Biol, 2011. **2011**: p. 451676.
22. Sternlicht, M.D., et al., *The stromal proteinase MMP3/stromelysin-1 promotes mammary carcinogenesis.* Cell, 1999. **98**(2): p. 137-46.
23. Hojilla, C.V., F.F. Mohammed, and R. Khokha, *Matrix metalloproteinases and their tissue inhibitors direct cell fate during cancer development*, in *Br J Cancer*. 2003. p. 1817-21.
24. Li, Y., W.P. Hively, and H.E. Varmus, *Use of MMTV-Wnt-1 transgenic mice for studying the genetic basis of breast cancer.* Oncogene, 2000. **19**(8): p. 1002-9.
25. Sundvall, M., et al., *Role of ErbB4 in breast cancer.* J Mammary Gland Biol Neoplasia, 2008. **13**(2): p. 259-68.
26. Turashvili G., B.J., Burkadze G., Kolar Z., *Mammary gland development and cancer.* Čes.-slov. Patol., 2005. **41**(3): p. 94-101.
27. Garcia-Martinez, J.M., et al., *A non-catalytic function of the Src family tyrosine kinases controls prolactin-induced Jak2 signaling.* Cell Signal, 2010. **22**(3): p. 415-26.
28. Peeva, E., et al., *Prolactin modulates the naive B cell repertoire.* J Clin Invest, 2003. **111**(2): p. 275-83.
29. Bridges, R.S., et al., *Endocrine communication between conceptus and mother: placental lactogen stimulation of maternal behavior.* Neuroendocrinology, 1996. **64**(1): p. 57-64.
30. Shingo, T., et al., *Pregnancy-Stimulated Neurogenesis in the Adult Female Forebrain Mediated by Prolactin.* Science, 2003. **299**(5603): p. 117-120.
31. Freeman, M.E., et al., *Prolactin: structure, function, and regulation of secretion.* Physiol Rev, 2000. **80**(4): p. 1523-631.
32. Bernard, V., et al., *New insights in prolactin: pathological implications.* Nat Rev Endocrinol, 2015. **11**(5): p. 265-75.
33. Cohen, L.E., F.E. Wondisford, and S. Radovick, *Role of Pit-1 in the gene expression of growth hormone, prolactin, and thyrotropin.* Endocrinol Metab Clin North Am, 1996. **25**(3): p. 523-40.
34. Bole-Feysot, C., et al., *Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice.* Endocr Rev, 1998. **19**(3): p. 225-68.
35. Leong, D.A., L.S. Frawley, and J.D. Neill, *Neuroendocrine control of prolactin secretion.* Annu Rev Physiol, 1983. **45**: p. 109-27.
36. Bose, S., et al., *A Pit-1 Binding Site Adjacent to E-box133 in the Rat PRL Promoter is Necessary for Pulsatile Gene Expression Activity.* Neurochem Res, 2016.

37. Devi, Y.S. and J. Halperin, *Reproductive actions of prolactin mediated through short and long receptor isoforms.* Mol Cell Endocrinol, 2014. **382**(1): p. 400-10.
38. Dagil, R., et al., *The WSXWS motif in cytokine receptors is a molecular switch involved in receptor activation: insight from structures of the prolactin receptor.* Structure, 2012. **20**(2): p. 270-82.
39. Goffin, V., et al., *Prolactin: a hormone at the crossroads of neuroimmunoendocrinology.* Ann N Y Acad Sci, 1998. **840**: p. 498-509.
40. Liu, X., et al., *Stat5a is mandatory for adult mammary gland development and lactogenesis.* Genes Dev, 1997. **11**(2): p. 179-86.
41. Basham, B., et al., *In vivo identification of novel STAT5 target genes.* Nucleic Acids Res, 2008. **36**(11): p. 3802-18.
42. Barash, I., A. Institute of Animal Science, The Volcani Center, Bet- Dagan, Israel, and A. Institute of Animal Science, The Volcani Center, P.O. Box 6, Bet- Dagan 50250, Israel., *Stat5 in the mammary gland: Controlling normal development and cancer.* Journal of Cellular Physiology, 2016. **209**(2): p. 305-313.
43. Chida, D., et al., *Transcriptional regulation of the beta-casein gene by cytokines: cross-talk between STAT5 and other signaling molecules.* Mol Endocrinol, 1998. **12**(11): p. 1792-806.
44. Song, G., G. Ouyang, and S. Bao, *The activation of Akt/PKB signaling pathway and cell survival.* J Cell Mol Med, 2005. **9**(1): p. 59-71.
45. Hankinson, S.E., et al., *Plasma prolactin levels and subsequent risk of breast cancer in postmenopausal women.* J Natl Cancer Inst, 1999. **91**(7): p. 629-34.
46. Clevenger, C.V., S.L. Gadd, and J. Zheng, *New mechanisms for PRLr action in breast cancer.* Trends Endocrinol Metab, 2009. **20**(5): p. 223-9.
47. Gadd, S.L. and C.V. Clevenger, *Ligand-independent dimerization of the human prolactin receptor isoforms: functional implications.* Mol Endocrinol, 2006. **20**(11): p. 2734-46.
48. Bogorad, R.L., et al., *Identification of a gain-of-function mutation of the prolactin receptor in women with benign breast tumors.* PNAS, 2008. **105**(38): p. 14533–14538.
49. George, D. *Prolactin & Prolactin Signalnig Pathways in Cancer Risk & Tumorigenesis*. 2016 [cited 2016 14.04.2016]; Available from: http://scrippslabs.com/scripps-news-volume-16-number-1/.
50. Brooks, C.L., *Molecular mechanisms of prolactin and its receptor.* Endocr Rev, 2012. **33**(4): p. 504-25.
51. Binart, N., A. Bachelot, and J. Bouilly, *Impact of prolactin receptor isoforms on reproduction.* Trends Endocrinol Metab, 2010. **21**(6): p. 362-8.
52. Clapp, C., et al., *The 16-kilodalton N-terminal fragment of human prolactin is a potent inhibitor of angiogenesis.* Endocrinology, 1993. **133**(3): p. 1292-9.
53. Salomon, D.S., C. Bianco, and M. De Santis, *Cripto: a novel epidermal growth factor (EGF)-related peptide in mammary gland development and neoplasia.* Bioessays, 1999. **21**(1): p. 61-70.
54. Masuda, H., et al., *Role of Epidermal Growth Factor Receptor in Breast Cancer.* Breast Cancer Res Treat, 2012. **136**(2).

55. Wang, Y. and B.P. Zhou, *Epithelial-mesenchymal transition in breast cancer progression and metastasis.* Chin J Cancer, 2011. **30**(9): p. 603-11.
56. Morgillo, F., et al., *Heterodimerization of Insulin-like Growth Factor Receptor/Epidermal Growth Factor Receptor and Induction of Survivin Expression Counteract the Antitumor Action of Erlotinib.* 2006.
57. Slamon, D.J., et al., *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2.* N Engl J Med, 2001. **344**(11): p. 783-92.
58. Hudis, C.A., *Trastuzumab — Mechanism of Action and Use in Clinical Practice.* N Engl J Med, 2009. **357**: p. 39-51.
59. Wiesen, J.F., et al., *Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development.* Development, 1999. **126**(2): p. 335-44.
60. Clark, D.E., et al., *ERBB4/HER4 potentiates STAT5A transcriptional activity by regulating novel STAT5A serine phosphorylation events.* J Biol Chem, 2005. **280**(25): p. 24175-80.
61. Muraoka-Cook, R.S., et al., *The Intracellular Domain of ErbB4 Induces Differentiation of Mammary Epithelial Cells.* Mol Biol Cell, 2006. **17**(9): p. 4118-29.
62. Walden, P.D., et al., *Evidence that the mammary fat pad mediates the action of growth hormone in mammary gland development.* Endocrinology, 1998. **139**(2): p. 659-62.
63. Chatterjee, S., et al., *Loss of Igfbp7 causes precocious involution in lactating mouse mammary gland.* PLoS One, 2014. **9**(2): p. e87858.
64. Richards, R.G., et al., *Mammary gland branching morphogenesis is diminished in mice with a deficiency of insulin-like growth factor-I (IGF-I), but not in mice with a liver-specific deletion of IGF-I.* Endocrinology, 2004. **145**(7): p. 3106-10.
65. Christopoulos, P.F., P. Msaouel, and M. Koutsilieris, *The role of the insulin-like growth factor-1 system in breast cancer.* Molecular Cancer, 2015. **14**(1): p. 1.
66. Kaaks, R., et al., *Insulin-like growth factor I and risk of breast cancer by age and hormone receptor status-A prospective study within the EPIC cohort.* Int J Cancer, 2014. **134**(11): p. 2683-90.
67. Hiraga, T., et al., *Bone-derived IGF mediates crosstalk between bone and breast cancer cells in bony metastases.* Cancer Res, 2012. **72**(16): p. 4238-49.
68. Toriola, A.T., et al., *Circulating insulin-like growth factor-I in pregnancy and maternal risk of breast cancer.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(8): p. 1798-801.
69. Sureshbabu, A., et al., *IGFBP5 induces cell adhesion, increases cell survival and inhibits cell migration in MCF-7 human breast cancer cells.* J Cell Sci, 2012. **125**(Pt 7): p. 1693-705.
70. Rocha, R.L., et al., *Correlation of insulin-like growth factor-binding protein-3 messenger RNA with protein expression in primary breast cancer tissues: detection of higher levels in tumors with poor prognostic features.* J Natl Cancer Inst, 1996. **88**(9): p. 601-6.

71. Vestey, S.B., et al., *Immunohistochemical expression of insulin-like growth factor binding protein-3 in invasive breast cancers and ductal carcinoma in situ: implications for clinicopathology and patient outcome.* Breast Cancer Res, 2005. **7**(1): p. R119-29.
72. Kleinberg, D.L., *Role of IGF-I in normal mammary development.* Breast Cancer Res Treat, 1998. **47**(3): p. 201-8.
73. Brisken, C. and B. O'Malley, *Hormone Action in the Mammary Gland.* Cold Spring Harb Perspect Biol, 2010. **2**(12).
74. Jordan, V.C., *Molecular mechanisms of antiestrogen action in breast cancer.* Breast Cancer Res Treat, 1994. **31**(1): p. 41-52.
75. Kumar, V. and P. Chambon, *The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer.* Cell, 1988. **55**(1): p. 145-56.
76. Lee, Y. and J. Gorski, *Estrogen-induced transcription of the progesterone receptor gene does not parallel estrogen receptor occupancy.* Proc Natl Acad Sci U S A, 1996. **93**(26): p. 15180-4.
77. Gudelsky, G.A., D.D. Nansel, and J.C. Porter, *Role of estrogen in the dopaminergic control of prolactin secretion.* Endocrinology, 1981. **108**(2): p. 440-4.
78. Sommer, S. and S.A. Fuqua, *Estrogen receptor and breast cancer.* Semin Cancer Biol, 2001. **11**(5): p. 339-52.
79. Maehle, B.O., et al., *Estrogen receptor beta--an independent prognostic marker in estrogen receptor alpha and progesterone receptor-positive breast cancer?* Apmis, 2009. **117**(9): p. 644-50.
80. Jr, R.C.B., et al., *2000 Update of Recommendations for the Use of Tumor Markers in Breast and Colorectal Cancer: Clinical Practice Guidelines of the American Society of Clinical Oncology*.* Journal of Clinical Oncology, 2001. **19**(6): p. 1865-1878.
81. Jordan, V.C., *Tamoxifen: catalyst for the change to targeted therapy.* Eur J Cancer, 2008. **44**(1): p. 30-8.
82. Osborne, C.K., *Tamoxifen in the treatment of breast cancer.* N Engl J Med, 1998. **339**(22): p. 1609-18.
83. Onitilo, A.A., et al., *Breast Cancer Subtypes Based on ER/PR and Her2 Expression: Comparison of Clinicopathologic Features and Survival.* Clin Med Res, 2009. **7**(1-2): p. 4-13.
84. Chan, S., E.A. Choi, and Y. Shi, *Pre-mRNA 3'-end processing complex assembly and function.* Wiley Interdiscip Rev RNA, 2011. **2**(3): p. 321-35.
85. Shi, Y., *Alternative polyadenylation: New insights from global analyses.* RNA, 2012. **18**(12): p. 2105-17.
86. Ağuş, H.H. and A.E.E. Bensan, *Mechanisms of mRNA polyadenylation.* Turkish Journal of Biology, 2015. **40**.
87. Tian, B., et al., *A large-scale analysis of mRNA polyadenylation of human and mouse genes.* Nucleic Acids Res, 2005. **33**(1): p. 201-12.
88. Erson-Bensan, A.E. and T. Can, *Alternative Polyadenylation: Another foe in cancer.* Mol Cancer Res, 2016.
89. Colgan, D.F. and J.L. Manley, *Mechanism and regulation of mRNA polyadenylation.* Genes & Dev., 1997. **11**: p. 2755-2766.

90. de Vries, H., et al., *Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors.* Embo j, 2000. **19**(21): p. 5895-904.
91. Derti, A., et al., *A quantitative atlas of polyadenylation in five mammals.* Genome Res, 2012. **22**(6): p. 1173-83.
92. Lutz, C.S. and A. Moreira, *Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression.* Wiley Interdiscip Rev RNA, 2011. **2**(1): p. 23-31.
93. Akman, H.B. and A.E. Erson-Bensan, *Alternative polyadenylation and its impact on cellular processes.* Microrna, 2014. **3**(1): p. 2-9.
94. Sandberg, R., et al., *Proliferating cells express mRNAs with shortened 3' UTRs and fewer microRNA target sites#.* Science, 2008. **320**(5883): p. 1643-7.
95. Mayr, C. and D.P. Bartel, *Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells.* Cell, 2009. **138**(4): p. 673.
96. Ji, Z., et al., *Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 7028-33.
97. Watson, C. *Department of Pathology: Mammary Apoptosis and Development Group*. 2016 [cited 2016 4 February]; Available from: http://old-www.path.cam.ac.uk/%7Emadgroup/microarraysummary.shtml.
98. Clarkson, R.W. and C.J. Watson, *Microarray analysis of the involution switch.* J Mammary Gland Biol Neoplasia, 2003. **8**(3): p. 309-19.
99. Akman, B.H., T. Can, and A.E. Erson-Bensan, *Estrogen-induced upregulation and 3'-UTR shortening of CDC6.* Nucleic Acids Res, 2012. **40**(21): p. 10679-88.
100. Saeed, A.I., et al., *TM4: a free, open-source system for microarray data management and analysis.* Biotechniques, 2003. **34**(2): p. 374-8.
101. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
102. Huang, D.W., et al., *The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists.* Genome Biol, 2007. **8**(9): p. R183.
103. Ball, R.K., et al., *Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line.* Embo j, 1988. **7**(7): p. 2089-95.
104. Bustin, S.A., et al., *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments.* Clin Chem, 2009. **55**(4): p. 611-22.
105. Tallkvist, J., et al., *A model of secreting murine mammary epithelial HC11 cells comprising endogenous Bcrp/Abcg2 expression and function.* Cell Biol Toxicol, 2015. **31**(2): p. 111-20.
106. Fleige, S., et al., *Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR.* Biotechnol Lett, 2006. **28**(19): p. 1601-13.

107. Clarkson, R.W., et al., *Gene expression profiling of mammary gland development reveals putative roles for death receptors and immune mediators in post-lactational regression.* Breast Cancer Res, 2004. **6**(2): p. R92-109.
108. Konno, H., et al., *Acquired xanthine dehydrogenase expression shortens survival in patients with resected adenocarcinoma of lung.* Tumour Biol, 2012. **33**(5): p. 1727-32.
109. Nishino, T., *The Conversion of Xanthine Dehydrogenase to Xanthine Oxidase and the Role of the Enzyme in Reperfusion Injury.* The Journal of Biochemistry, 1994. **116**(1): p. 1-6.
110. Kurosaki, M., et al., *Expression of xanthine oxidoreductase in mouse mammary epithelium during pregnancy and lactation: regulation of gene expression by glucocorticoids and prolactin.* Biochem J, 1996. **319 (Pt 3)**: p. 801-10.
111. Vorbach, C., A. Scriven, and M.R. Capecchi, *The housekeeping gene xanthine oxidoreductase is necessary for milk fat droplet enveloping and secretion: gene sharing in the lactating mammary gland.* Genes Dev, 2002. **16**(24): p. 3223-35.
112. Bionaz, M. and J.J. Loor, *Gene networks driving bovine milk fat synthesis during the lactation cycle.* BMC Genomics, 2008. **9**: p. 366.
113. Cook, W., et al., *Differential immunohistochemical localization of xanthine oxidase in normal and neoplastic human breast epithelium.* Int J Oncol, 1997. **11**(5): p. 1013-7.
114. Linder, N., et al., *Down-regulated xanthine oxidoreductase is a feature of aggressive breast cancer.* Clin Cancer Res, 2005. **11**(12): p. 4372-81.
115. Vicente, C.M., et al., *Syndecan-2 is upregulated in colorectal cancer cells through interactions with extracellular matrix produced by stromal fibroblasts.* BMC Cell Biology, 2013. **14**(1): p. 1.
116. Lim, H.C., H.A. Multhaupt, and J.R. Couchman, *Cell surface heparan sulfate proteoglycans control adhesion and invasion of breast carcinoma cells.* Mol Cancer, 2015. **14**(1).
117. Lim, H.C. and J.R. Couchman, *Syndecan-2 regulation of morphology in breast carcinoma cells is dependent on RhoGTPases.* Biochim Biophys Acta, 2014. **1840**(8): p. 2482-90.
118. Modrowski, D., et al., *Syndecan-2 overexpression induces osteosarcoma cell apoptosis: Implication of syndecan-2 cytoplasmic domain and JNK signaling.* Bone, 2005. **37**(2): p. 180-9.
119. Yao, D.W., et al., *Thyroid hormone responsive (THRSP) promotes the synthesis of medium-chain fatty acids in goat mammary epithelial cells.* J Dairy Sci, 2016. **99**(4): p. 3124-33.
120. Harvatine, K.J. and D.E. Bauman, *SREBP1 and thyroid hormone responsive spot 14 (S14) are involved in the regulation of bovine mammary lipid synthesis during diet-induced milk fat depression and treatment with CLA.* J Nutr, 2006. **136**(10): p. 2468-74.
121. Kinlaw, W.B., et al., *Spot 14: A marker of aggressive breast cancer and a potential therapeutic target.* Endocrinology, 2006. **147**(9): p. 4048-55.
122. Menendez, J.A. and R. Lupu, *Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis.* Nat Rev Cancer, 2007. **7**(10): p. 763-77.

123. Burger, A.M., et al., *Essential roles of IGFBP-3 and IGFBP-rP1 in breast cancer.* Eur J Cancer, 2005. **41**(11): p. 1515-27.
124. Benatar, T., et al., *IGFBP7 reduces breast tumor growth by induction of senescence and apoptosis pathways.* Breast Cancer Res Treat, 2012. **133**(2): p. 563-73.
125. Amemiya, Y., et al., *Insulin like growth factor binding protein-7 reduces growth of human breast cancer cells and xenografted tumors.* Breast Cancer Res Treat, 2011. **126**(2): p. 373-84.
126. Wajapeyee, N., et al., *Oncogenic BRAF induces senescence and apoptosis through pathways mediated by the secreted protein IGFBP7.* Cell, 2008. **132**(3): p. 363-74.
127. Zhou, Z.J., et al., *Overexpression of HnRNP A1 promotes tumor invasion through regulating CD44v6 and indicates poor prognosis for hepatocellular carcinoma.* Int J Cancer, 2013. **132**(5): p. 1080-9.
128. Lemieux, B., et al., *A Function for the hnRNP A1/A2 Proteins in Transcription Elongation.* PLoS One, 2015. **10**(5): p. e0126654.
129. Mayeda, A. and A.R. Krainer, *Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2.* Cell, 1992. **68**(2): p. 365-75.
130. Ushigome, M., et al., *Up-regulation of hnRNP A1 gene in sporadic human colorectal cancers.* Int J Oncol, 2005. **26**(3): p. 635-40.
131. Yu, C., et al., *Oral squamous cancer cell exploits hnRNP A1 to regulate cell cycle and proliferation.* J Cell Physiol, 2015. **230**(9): p. 2252-61.
132. Loh, T.J., et al., *CD44 alternative splicing and hnRNP A1 expression are associated with the metastasis of breast cancer.* Oncol Rep, 2015. **34**(3): p. 1231-8.
133. Jean-Philippe, J., S. Paz, and M. Caputi, *hnRNP A1: The Swiss Army Knife of Gene Expression.* International Journal of Molecular Sciences, 2013. **14**(9): p. 18999-19024.
134. Fillmore, C. and C. Kuperwasser, *Human breast cancer stem cell markers CD44 and CD24: enriching for cells with functional properties in mice or in man?* Breast Cancer Res, 2007. **9**(3): p. 303.
135. and, K.L.K. and H.J. Yost, *Heparan Sulfate Core Proteins in Cell-Cell Signaling.* http://dx.doi.org/10.1146/annurev.genet.37.061103.090226, 2003. **37**: p. 461-484.
136. David, C.J., et al., *HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer.* Nature, 2010. **463**(7279): p. 364-8.
137. Akman, H.B., et al., *3'UTR shortening and EGF signaling: implications for breast cancer.* Hum Mol Genet, 2015. **24**(24): p. 6910-20.
138. Masamha, C.P., et al., *CFIm25 links alternative polyadenylation to glioblastoma tumour suppression.* Nature, 2014. **510**(7505): p. 412-6.

APPENDIX A

SIGNIFICANTLY 3'UTR ALTERED GENES AS A RESULT OF APADETECT AND SAM

Table A.1. Significantly 3'UTR shortened genes in gestation, lactation and involution timepoints compared to virgin animal data.

Gene Name	Probeset ID	Expected Score	Observed Score	Numerator	Denumerator	Fold Change (unlogged)
Thrsp	160306_at	-0.08743933	2.8764484	1.5354166	0.53378904	2.8990805
Sdc2	95104_at	0.34219623	2.4402165	1.6777265	0.6875318	3.1564372
Tomm70a	95431_at	0.4027095	2.113835	1.0322328	0.4883223	2.0460777
Tomm70a	95431_at	0.40351665	2.113835	1.0322328	0.4883223	2.0460777
Tomm70a	95431_at	0.4012922	2.113835	1.0322328	0.4883223	2.0460777
Dynlrb1	94862_i_at	-0.4389576	2.6819255	1.357039	0.50599426	2.5614488
Calr	93523_at	-0.028737796	2.215306	1.2934759	0.5838813	2.4533284
Tomm70a	95431_at	0.42596865	2.167821	1.3102341	0.6044014	2.4886553
Tomm70a	95431_at	0.425634	2.167821	1.3102341	0.6044014	2.4886553
Tomm70a	95431_at	0.4262605	2.167821	1.3102341	0.6044014	2.4886553
Msr1	94140_at	-0.044493698	2.1216543	1.3954065	0.6576974	2.6497717
P4hb	93108_at	0.19894336	2.1073477	1.4657496	0.69554234	2.7788632
P4hb	93108_at	0.19936734	2.1073477	1.4657496	0.69554234	2.7788632
Atp5f1	96898_at	-0.33518565	2.0699317	1.1939492	0.57680607	2.292174
Agpat6	100996_at	-0.046527408	1.9426229	1.2356763	0.6360865	2.3543055
Chaf1a	95527_at	0.588796	1.9422514	1.103032	0.5679141	2.1470962
Srpr	103680_at	0.008109565	1.9209867	1.0457983	0.54440683	2.0660815
Asf1a	98914_at	0.09437716	1.8928379	1.1210711	0.59227	2.1725223
Asf1a	98914_at	0.09467728	1.8928379	1.1210711	0.59227	2.1725223
Asf1a	98914_at	0.0953949	1.8928379	1.1210711	0.59227	2.1725223
Asf1a	98914_at	0.09513758	1.8928379	1.1210711	0.59227	2.1725223
Araf	102286_at	0.05671482	1.8723367	0.94285095	0.5035691	1.922523
Ap3d1	96185_at	0.10027873	1.8679647	0.9566512	0.51213557	1.9409732
Syt6	100364_at	-0.41537333	1.8098857	0.9343582	0.5162526	1.9115487
Syt6	100364_at	-0.41449246	1.8098857	0.9343582	0.5162526	1.9115487
Mphosph10	104022_at	-0.097441815	1.7861207	1.2320898	0.6898133	2.348133
D10Ertd32 2e	96355_at	0.107118815	1.7651103	1.2765334	0.7232032	2.4467862
Myl9	96939_at	-0.43595788	1.7285092	1.4494002	0.8385262	2.7860203
Cyp4a10	92600_f_at	-0.26490086	1.7276773	0.8930257	0.5168938	1.8575413
Ap2s1	96638_at	-0.12008268	1.6948189	1.2953911	0.7643242	2.462841

Table A.1. (continued)

Gene Name	Probeset ID	Expected Score	Observed Score	Numerator	Denumerator	Fold Change (unlogged)
Osbp	103348_at	0.9653401	1.6895862	1.053608	0.62358934	2.072272
Osbp	103348_at	0.96462405	1.6895862	1.053608	0.62358934	2.072272
Golga7	93256_at	-0.04620275	1.6768619	1.0979421	0.65476	2.134411
Hras1	160536_at	-0.06890363	1.6597272	1.3171616	0.7936012	2.4426546
Rnpep	92865_at	0.42730093	1.6563503	0.8435519	0.5092835	1.7942865
Glb1l	94480_at	-0.75897974	1.6250733	1.047857	0.644806	2.0790906
Tbx2	104655_at	0.17689271	1.6046137	1.1193689	0.69759405	2.1744864
Tbx2	104655_at	0.17633086	1.6046137	1.1193689	0.69759405	2.1744864
Map2k1ip1	160126_at	-0.3153747	1.5942353	1.0090749	0.63295233	2.0167267
Rgn	93766_at	0.056296993	1.5839921	1.109369	0.70036274	2.1796184
Acly	160207_at	0.18857776	1.5803635	0.9870758	0.62458783	1.9741149
Acly	160207_at	0.18814145	1.5803635	0.9870758	0.62458783	1.9741149
Aco1	100147_at	-0.3018387	1.5777057	1.4020804	0.8886831	2.6510625
Vamp4	102317_at	-0.654789	1.5728023	0.8645457	0.54968494	1.8190627
Vamp4	102317_at	-0.65692383	1.5728023	0.8645457	0.54968494	1.8190627
Slc30a5	94471_r_at	0.26063755	1.5671767	1.0679504	0.6814486	2.1096394
Paqr7	99142_at	-0.25085568	1.5641829	0.8641335	0.5524504	1.8206263
Paqr7	99142_at	-0.25115526	1.5641829	0.8641335	0.5524504	1.8206263
Myg1	95651_at	0.41967124	1.5586994	0.8899112	0.5709319	1.8512714
Stard7	99471_at	-0.45674005	1.5557324	1.5056934	0.9678358	2.9131284
Elovl3	103469_at	1.3156911	1.5539975	1.0637546	0.6845278	2.0759435
Elovl3	103469_at	1.2707732	1.5539975	1.0637546	0.6845278	2.0759435
Rpe	99647_at	-0.7904179	1.5269954	1.2533274	0.82078004	2.3319798
Mtap7	93729_at	0.07301967	1.5191464	0.81092536	0.5338033	1.7540702
Mtap7	93729_at	0.0743608	1.5191464	0.81092536	0.5338033	1.7540702
Mtap7	93729_at	0.074444525	1.5191464	0.81092536	0.5338033	1.7540702
Mtap7	93729_at	0.073857434	1.5191464	0.81092536	0.5338033	1.7540702
Nme1	92794_f_at	0.17948818	1.5128051	1.2232044	0.80856705	2.2987869
Impa2	98420_at	0.7557634	1.5118734	0.87365603	0.5778632	1.8283235
Actc1	101029_f_at	-0.5689002	3.3405385	1.2087083	0.36183038	2.3128583
Mkrn1	101068_at	-0.19564249	2.9835377	0.99590707	0.33380073	1.9941275
Thrsp	160306_at	-0.14559242	6.377342	1.5167996	0.23784196	2.8630562
Msr1	94140_at	-0.07631041	6.5144587	1.6101015	0.24715814	3.0543122
Thrsp	160306_at	-0.14762194	6.3267946	1.8530921	0.29289588	3.614477
P4hb	93108_at	0.29853848	5.689771	1.7460037	0.30686712	3.347944
P4hb	93108_at	0.29930574	5.689771	1.7460037	0.30686712	3.347944
Sdc2	95104_at	0.58209574	5.441344	2.3749518	0.43646422	5.116561
Vps26b	99139_at	0.001837913	5.030506	1.1071022	0.2200777	2.1542566
Vps26b	99139_at	8.50E-04	5.030506	1.1071022	0.2200777	2.1542566
Otc	94414_at	0.058641244	4.978474	1.3673104	0.27464446	2.5802412

Table A.1. (continued)

Gene Name	Probeset ID	Expected Score	Observed Score	Numerator	Denumerator	Fold Change (unlogged)
Col6a1	95493_at	0.13884686	4.5315194	1.5525588	0.34261328	2.9443412
Mid1ip1	95134_at	0.07839476	4.4783826	1.2414038	0.27719915	2.3606632
Tomm70a	95431_at	0.6883265	4.376637	1.0541981	0.24086945	2.0775342
Tomm70a	95431_at	0.687014	4.376637	1.0541981	0.24086945	2.0775342
Tomm70a	95431_at	0.6883265	4.376637	1.0541981	0.24086945	2.0775342
Col6a1	95493_at	0.13985212	4.2744875	1.5039363	0.35184014	2.8461206
A630038E17Rik	92328_at	0.5473876	4.249613	0.9658132	0.22727087	1.9527959
Lst1	103571_at	0.9292148	3.8738043	1.0685918	0.27585074	2.100118
Sdc2	95104_at	0.3971197	4.021124	2.2713633	0.56485784	4.7620387
Ubap2l	96578_r_at	-0.4060777	3.4155314	1.5876622	0.4648361	2.9932334
Ubap2l	96578_r_at	-0.40529603	3.4155314	1.5876622	0.4648361	2.9932334
Ubap2l	96578_r_at	-0.40676937	3.4155314	1.5876622	0.4648361	2.9932334

Table A.2. Significantly 3'UTR lengthened genes in gestation, lactation and involution timepoints compared to virgin animal data.

Gene Name	Probeset ID	Expected Score	Observed Score	Numerator	Denumerator	Fold Change (unlogged)
Ccdc80	160298_at	0.7347165	-6.0042467	-2.0234318	0.3370001	0.24481522
Ccdc80	160298_at	0.7316647	-6.0042467	-2.0234318	0.3370001	0.24481522
Fn1	92852_at	-1.4722458	-5.560309	-2.0987124	0.37744528	0.23299904
C3	93497_at	1.0193791	-4.7565665	-2.1686382	0.45592514	0.22044314
Dcp1a	96223_at	0.40100068	-3.1364796	-1.302282	0.41520497	0.4042858
Igfbp7	160527_at	-0.35146722	-6.2122526	-1.2973924	0.20884411	0.40680426
Ccdc80	160298_at	0.83119005	-6.985122	-2.1040077	0.30121273	0.23207775
Ccdc80	160298_at	0.8334132	-6.985122	-2.1040077	0.30121273	0.23207775
Ccl6	92849_at	0.2878296	-7.1367683	-1.3500601	0.18916968	0.39256513
Fn1	92852_at	-1.5151027	-8.184168	-2.2350612	0.27309573	0.21138708
Amy1	101058_at	-0.5736126	-10.101186	-1.473319	0.14585605	0.36017635
Amy1	101058_at	-0.57249266	-10.101186	-1.473319	0.14585605	0.36017635
Amy1	101058_at	-0.5716764	-10.101186	-1.473319	0.14585605	0.36017635
Cd302	160387_at	-0.57734084	-2.437032	-0.93831307	0.38502288	0.52155864
Cd302	160387_at	-0.5757644	-2.437032	-0.93831307	0.38502288	0.52155864
Slc6a6	94405_at	-0.15565088	-2.4654264	-1.0437719	0.42336363	0.48491493
Prx	102214_at	-0.117079556	-2.5181634	-0.8460578	0.33598208	0.5563085
Prx	102214_at	-0.11712727	-2.5181634	-0.8460578	0.33598208	0.5563085
1700022I11 Rik	100710_at	-0.32282263	-2.5850296	-1.2790742	0.4948006	0.41014183
Rbbp4	92647_at	-0.27881104	-2.6542656	-1.0726054	0.40410626	0.47477308
Cisd1	94526_at	0.08943138	-2.67751	-1.0870221	0.40598243	0.47015202
Ccl6	92849_at	0.17413726	-2.8493695	-1.0510076	0.3688562	0.4827846
Ndufab1	96909_at	-0.136223	-2.876703	-1.2353354	0.42942747	0.4242529
Serpinf1	93574_at	0.12224528	-2.9247985	-1.2060542	0.41235465	0.43417096
Serpinf1	93574_at	0.122018024	-2.9247985	-1.2060542	0.41235465	0.43417096
Dctn2	94300_f_at	0.10492484	-3.0689163	-1.0367501	0.33782285	0.48740703
Psmb4	98557_f_at	-0.39667904	-3.03445	-1.5575883	0.5133017	0.3392575
Dcp1a	96223_at	0.34801358	-2.7692049	-1.2358328	0.44627714	0.4234749
Psmb4	98557_f_at	-0.39589623	-3.03445	-1.5575883	0.5133017	0.3392575
Amy1	101058_at	-0.3053659	-3.033838	-1.3910086	0.45849797	0.38167515
Amy1	101058_at	-0.30460796	-3.033838	-1.3910086	0.45849797	0.38167515
Amy1	101058_at	-0.3040878	-3.033838	-1.3910086	0.45849797	0.38167515
6530401D1 7Rik	95523_at	0.062255632	-2.5035086	-1.3374332	0.53422356	0.3960899
6530401D1 7Rik	95523_at	0.06184444	-2.5035086	-1.3374332	0.53422356	0.3960899
Igfbp7	160527_at	-0.23236452	-2.9047282	-1.2476394	0.4295202	0.42170432
Amy1	101058_at	-0.3648927	-2.9711215	-1.3305286	0.44782034	0.39925072
Amy1	101058_at	-0.36614767	-2.9711215	-1.3305286	0.44782034	0.39925072
Amy1	101058_at	-0.3670534	-2.9711215	-1.3305286	0.44782034	0.39925072
Prx	102214_at	-0.10971035	-2.9769125	-0.96394706	0.32380766	0.51264876
Prx	102214_at	-0.11008149	-2.9769125	-0.96394706	0.32380766	0.51264876
Ccdc80	160298_at	0.4936258	-3.6962163	-1.7677268	0.47825307	0.29272932

Table A.2. (continued)

Gene Name	Probeset ID	Expected Score	Observed Score	Numerator	Denumerator	Fold Change (unlogged)
Prnp	100606_at	-0.5109327	-3.1108468	-1.047828	0.33683047	0.48368144
Amy1	101058_at	-0.36747986	-3.1295674	-1.3674788	0.4369546	0.38872954
Amy1	101058_at	-0.36684224	-3.1295674	-1.3674788	0.4369546	0.38872954
Amy1	101058_at	-0.36787856	-3.1295674	-1.3674788	0.4369546	0.38872954
Igfbp4	94222_at	0.18627863	-3.2243187	-1.6974916	0.52646524	0.30763468
Igfbp4	94222_at	0.18673037	-3.2243187	-1.6974916	0.52646524	0.30763468
Igfbp4	94222_at	0.18710731	-3.2243187	-1.6974916	0.52646524	0.30763468
Dctn2	94300_f_at	0.10488444	-3.2390249	-1.1035984	0.3407193	0.46538076
C3	93497_at	0.6694693	-3.4350982	-1.9726727	0.57426965	0.25220528
Scp2	93278_at	-0.33409253	-3.4603062	-1.5385045	0.44461513	0.34464434
Scp2	93278_at	-0.33357954	-3.4603062	-1.5385045	0.44461513	0.34464434
6530401D17Rik	95523_at	0.06350401	-3.9034317	-1.8432592	0.47221506	0.2772509
6530401D17Rik	95523_at	0.06372064	-3.9034317	-1.8432592	0.47221506	0.2772509
Fn1	92852_at	-1.0117306	-4.379255	-2.0512135	0.46839327	0.24009731
Ccdc80	160298_at	0.5032539	-4.3891525	-2.094289	0.47715113	0.23327343
Ccdc80	160298_at	0.50164616	-4.3891525	-2.094289	0.47715113	0.23327343
Igfbp7	160527_at	-0.22313705	-4.6446695	-1.8490676	0.3981053	0.27745968
C3	93497_at	0.8791339	-3.9463212	-1.8647487	0.47252837	0.27212775
Prx	102214_at	-0.13691843	-4.295542	-0.98981345	0.23042808	0.50356895
Prx	102214_at	-0.13725594	-4.295542	-0.98981345	0.23042808	0.50356895
Scp2	93278_at	-0.40880486	-4.662318	-1.4634945	0.31389847	0.36267456
Scp2	93278_at	-0.40977666	-4.662318	-1.4634945	0.31389847	0.36267456
Amy1	101058_at	-0.4500557	-5.1038704	-1.4968965	0.29328653	0.35485592
Amy1	101058_at	-0.4513135	-5.1038704	-1.4968965	0.29328653	0.35485592
Amy1	101058_at	-0.45155838	-5.1038704	-1.4968965	0.29328653	0.35485592
6530401D17Rik	95523_at	0.095479526	-5.268626	-1.9057484	0.36171636	0.26552215
6530401D17Rik	95523_at	0.09605792	-5.268626	-1.9057484	0.36171636	0.26552215
Ccdc80	160298_at	0.65522313	-5.5121365	-2.1081247	0.38245147	0.23135093
Ccdc80	160298_at	0.65713245	-5.5121365	-2.1081247	0.38245147	0.23135093
Igfbp7	160527_at	-0.2799488	-6.2570386	-1.7115357	0.27353767	0.3050868
Arih1	95563_at	-0.007290845	-1.8647089	-0.85652816	0.45933613	0.5519452
Wdr45	96728_at	0.042745184	-1.890307	-0.8396152	0.44416872	0.5588589
Wdr45	96728_at	0.04291226	-1.890307	-0.8396152	0.44416872	0.5588589
Dcp1a	96223_at	0.29458287	-1.9293393	-1.0856674	0.5627146	0.4712549
Serpinf1	93574_at	0.0969018	-1.9797616	-1.0232552	0.5168578	0.4937018
Serpinf1	93574_at	0.09674117	-1.9797616	-1.0232552	0.5168578	0.4937018
AW112010	100944_at	0.9902285	-2.016299	-1.2295862	0.60982335	0.4217598
Sucla2	93502_r_at	0.31776065	-2.043036	-1.1475719	0.56169933	0.4488285
Adra1a	101740_at	0.3123566	-2.0760927	-1.1993774	0.57770896	0.4319903
Adra1a	101740_at	0.3134181	-2.0760927	-1.1993774	0.57770896	0.4319903
Epb4.1l2	101500_at	0.069969915	-2.2660863	-0.9895121	0.43666124	0.50344735

Table A.2. (continued)

Gene Name	Probeset ID	Expected Score	Observed Score	Numerator	Denumerator	Fold Change (unlogged)
Prx	102214_at	-0.09413991	-2.2961576	-1.0699054	0.4659547	0.47693133
Prx	102214_at	-0.0938224	-2.2961576	-1.0699054	0.4659547	0.47693133
Cyp2e1	93996_at	-0.06519679	-2.345378	-1.7419143	0.7427009	0.29155678
C3	93497_at	0.5670536	-2.3729365	-1.4872259	0.6267449	0.35241863
Scp2	93278_at	-0.27696177	-2.4378893	-1.1513114	0.47225744	0.44985914
Scp2	93278_at	-0.2764627	-2.4378893	-1.1513114	0.47225744	0.44985914
Ccdc80	160298_at	0.4197997	-2.7941122	-1.6823195	0.6020945	0.3118721
Ccdc80	160298_at	0.42058083	-2.7941122	-1.6823195	0.6020945	0.3118721
6530401D17Rik	95523_at	0.055144787	-3.008515	-1.660258	0.551853	0.31499234
6530401D17Rik	95523_at	0.05560436	-3.008515	-1.660258	0.551853	0.31499234